CARDIOVASCULAR SYSTEM MONITORING USING WEARABLE SENSORS

by

Arash Shokouhmand

A DISSERTATION

Submitted to the Faculty of the Stevens Institute of Technology
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Arash Shokouhmand, Candidate

ADVISORY COMMITTEE

| | |
|---|---|
| Dr. Negar Ebadi | Date |
| Dr. Yu-Dong Yao | Date |
| Dr. Hongbin Li | Date |
| Dr. Brendan Englot | Date |

STEVENS INSTITUTE OF TECHNOLOGY
Castle Point on Hudson
Hoboken, NJ 07030
2023

©2023, Arash Shokouhmand. All rights reserved.

CARDIOVASCULAR SYSTEM MONITORING USING WEARABLE SENSORS

## ABSTRACT

The objective of this dissertation is to introduce the utilization of wearable sensors to observe the cardiovascular system in both fetuses and individuals with cardiovascular complications. The first chapter of the dissertation reviews the clinical significance of hemodynamic monitoring, cardiovascular disease detection, and continuous fetal heart rate estimation, as well as the current monitoring techniques and their limitations.

The second chapter discusses the limitations of current methods for detecting valvular heart disease (VHD) and peripheral artery disease (PAD), and introduces wearable-based methodology to address the limitations. For this purpose, wearable inertial measurement units (IMU) including seismo-cardiogram (SCG), gyro-cardiogram (GCG), and accelerometer contact microphone (ACM) were utilized in combination with machine learning (ML) and deep learning (DL) techniques to identify abnormal patterns corresponding to conditions in central and peripheral arteries. Furthermore, this chapter introduces strain sensors for cuffless estimation of blood pressure at the fingertip. The method proposes a novel feature extraction method to estimate blood pressure (BP) from pulse signals acquired by a strain plethysmography (SPG) sensor.

The third chapter investigates the detection of abnormality in left ventricular ejection fraction (LVEF) and mean pressure gradient (MPG) by fusing information from biopotential and vibrational recordings on the chest wall. The proposed framework leverages the time intervals of the vibrational signals to classify the AS severity with high

accuracy. The chapter continues by presenting novel strategies for augmenting plethysmography methods for hemodynamic monitoring. To this end, the use of micro-electromechanical systems (MEMS) strain sensors is investigated for estimating hemodynamic parameters.

The fourth chapter presents a new framework that combines ECG and inertial sensors to monitor fetal heart rate (FHR). Additionally, a signal-to-signal translation technique based on a generative adversarial network (GAN) is introduced to cancel fetal movement in abdominal ECG recordings. The chapter also presents the dual-path source separation to extract fetal ECG from single-channel abdominal ECG recordings.

The final chapter of the dissertation provides a summary of the developed methodologies for cardiovascular monitoring, and concludes the dissertation.

Author: Arash Shokouhmand

Advisor: Dr. Negar Ebadi

Date: July 12, 2023

Department: Electrical & Computer Engineering

Degree: Doctor of Philosophy

*To my beloved mother, Ashraf Sarvari*

*and*

*In the loving memory of my father, Javad Shokouhmand*

# Acknowledgment

First and foremost, I would like to express my sincere gratitude to my advisor and mentor, Professor Negar Ebadi, for her invaluable guidance, unwavering support, and exceptional mentorship throughout my academic journey. Professor Ebadi played a pivotal role in shaping the direction and success of this research project. Her influence and spirit of persistence will undoubtedly continue to inspire me in my future endeavors.

I am also grateful to Professor Yu-Dong Yao, Professor Hongbin Li, and Professor Brendan Englot for their valuable time to serve on the committee. Their thoughtful feedback and insightful suggestions have significantly helped to further develop and refine my research.

I would like to express my sincere gratitude to Professor Farrokh Ayazi, the director of the Integrated MEMS Laboratory at Georgia Tech, for his technical insights and contributions that have opened new doors throughout the research. I am also thankful to our medical collaborators, Dr. Philip Green, Dr. Amisha Patel, Dr. Joseph Puma, and Dr. Samiha Khan from Sorin Medical P.C., and Dr. Bruce Young and Dr. Clarel Antoine from NYU Langone Health, without whom clinical trials would not have been possible. I would like to thank Dr. Behnood Gholami, the CEO of Autonomous Healthcare, for his invaluable guidance which has greatly influenced my path towards professional growth. I am grateful for having Dr. Amir Mirbeik, the current CEO of RadioSight, as a labmate; his exceptional camaraderie has made my experience truly enjoyable.

I would like to extend a special thank you to my brother, Ehsan Shokouhmand, and his wife, Saeideh Fouji, for their exceptional support and unwavering presence in my life.

Ehsan’s guidance, encouragement, and kindness have been instrumental in shaping the person I am today. He has always been there for me, offering a listening ear, wise advice, and a shoulder to lean on. I would also like to express my gratitude to my other beloved siblings, Shahab Shokouhmand, Shakoura Shokouhmand, and Shiva Shokouhmand, my dear bother-in-law, Omid Golchin, and my sweet nieces, Arghavan Shokouhmand, Rosha Shokouhmand, and Shaina Shokouhmand. Their love, support, and companionship have been a constant source of strength and joy in my life. Thank you for being there for me, for the cherished memories we have created together, and for being an everlasting source of love and support. I am forever grateful for each and every one of you.

Last but not least, I would like to express my deepest gratitude to my beloved mother, Ashraf Sarvari, and my late father, Javad Shokouhmand, whose unwavering love, encouragement, and sacrifices have been the driving force behind my academic pursuits. My mother's enduring support and belief in my abilities have been a constant source of motivation, providing me with the strength to overcome challenges and strive for excellence. Although my father is no longer with us, his memory and the values he instilled in me continue to guide my path. His wisdom, determination, and unwavering belief in my potential have shaped my character and fueled my ambition. I am forever grateful to both of them for their immeasurable impact on my life. *This dissertation serves as a tribute to their love and support.*

# Table of Contents

ABSTRACT ..... III

1 INTRODUCTION ..... 1

1.1 SIGNIFICANCE OF VALVULAR HEART DISEASE DETECTION, PERIPHERAL ARTERY DISEASE DETECTION, AND HEMODYNAMIC MONITORING ..... 2

1.2 SIGNIFICANCE OF CONTINUOUS FETAL HEART RATE MONITORING ..... 6

1.3 DISSERTATION OBJECTIVES ..... 8

2 DIAGNOSIS OF CARDIOVASCULAR DISEASES USING RECORDINGS OF CARDIAC ACTIVITIES ON CENTRAL AND PERIPHERAL SITES ..... 9

2.1 DIAGNOSIS OF AORTIC STENOSIS AND STRATIFICATION OF ITS SEVERITY LEVEL USING CARDIOMECHANICAL ACTIVITIES ON THE CHEST WALL ..... 9

*2.1.1 Motivation ..... 11*

*2.1.2 Methods ..... 12*

*2.1.3 Experimental Results and Discussion ..... 24*

2.2 DIAGNOSIS OF COEXISTING VALVULAR HEART DISEASES USING IMAGE-TO-SEQUENCE TRANSLATION OF CONTACT MICROPHONE RECORDINGS ..... 37

*2.2.1 Background on Heart Sounds and Abnormalities ..... 41*

*2.2.2 Experimental Setup and Methodology ..... 43*

*2.2.3 Experimental Results ..... 55*

*2.2.4 Discussion ..... 62*

2.3 DIAGNOSIS OF PERIPHERAL ARTERY DISEASE USING BACKFLOW ABNORMALITIES IN PROXIMAL RECORDINGS OF ACCELEROMETER CONTACT MICROPHONE (ACM)..... 67

*2.3.1 Background on Arterial Tree Dynamics ..... 68*

*2.3.2 Experimental Setup and Methodology ..... 69*

*2.3.3 Experimental Results ..... 82*

*2.3.4 Discussion ..... 89*

2.4 MEMS FINGERTIP STRAIN PLETHYSMOGRAPHY FOR CUFFLESS ESTIMATION OF BLOOD PRESSURE..... 94

*2.4.1 Arterial Dilation in Fingertip ..... 98*

*2.4.2 Experimental Setup and Methodology ..... 99*

*2.4.3 Experimental Results ..... 109*

*2.4.4 Discussion ..... 119*

3 HEMODYNAMIC MONITORING USING WEARABLE SENSORS ..... 126

3.1 MEAN PRESSURE GRADIENT PREDICTION BASED ON CHEST ANGULAR MOVEMENTS AND HEART RATE VARIABILITY PARAMETERS..... 126

*3.1.1 Methodology for Classification of Mean Pressure Gradient..... 127*

*3.1.2 Experimental Results on MPG Classification ..... 132*

3.2 DETECTION OF LEFT VENTRICULAR EJECTION FRACTION USING FUSION OF ACCELEROMETER CONTACT MICROPHONE AND BIOPOTENTIAL SENSORS ..... 135

*3.2.1 Experimental Setup & Methodology..... 136*

*3.2.2 Experimental Results and Discussion..... 140*

3.3 FINGERTIP STRAIN PLETHYSMOGRAPHY: REPRESENTATION OF PULSE INFORMATION BASED ON VASCULAR VIBRATION ..... 143

*3.3.1 Fingertip Strain Plethysmography Methodology ..... 145*

*3.3.2 Proof-of-Concept Results & Discussion ..... 153*

3.4 A COMPARATIVE STUDY OF HEART RATE VARIABILITY PARAMETERS ESTIMATED FROM STRAIN PLETHYSMOGRAPHY RECORDINGS OF RADIAL AND FINGER ARTERIES ..... 155

*3.4.1 Experimental Setup and Methodology ..... 157*

*3.4.2 Experimental Results and Discussion ..... 161*

3.5 PROSPECTS OF CUFFLESS PULSE PRESSURE ESTIMATION FROM A CHEST-WORN ACCELEROMETER CONTACT MICROPHONE ..... 165

*3.5.1 Experimental Setup and Methodology ..... 166*

*3.5.2 Experimental Results and Discussion ..... 171*

4 FETAL HEART RATE ESTIMATION USING VIBRATIONAL AND BIOPOTENTIAL SENSORS ..... 175

4.1 FUSION OF BIOPOTENTIAL AND VIBRATIONAL SENSORS FOR FETAL HEART RATE ESTIMATION ..... 175

*4.1.1 Methodology ..... 177*

*4.1.2 Experimental Results and Discussion ..... 182*

4.2 FETAL MOVEMENT CANCELLATION IN ABDOMINAL ECG ELECTROCARDIOGRAM RECORDINGS USING SIGNAL-TO-SIGNAL TRANSLATION ..... 185

4.3 METHODOLOGY ..... 187

*4.3.1 Dataset ..... 188*

*4.3.2 Data Pre-Processing ..... 188*

*4.3.3 Signal-to-Signal Translation Network ..... 189*

*4.3.4 Training Procedure ..... 193*

4.4 EXPERIMENTAL RESULTS AND DISCUSSION ..... 194

*4.4.1 Evaluation Metrics ..... 194*

*4.4.2 Performance Evaluation and Discussion ..... 195*

4.5 FETAL ELECTROCARDIOGRAM EXTRACTION USING DUAL-PATH SOURCE SEPARATION OF SINGLE-CHANNEL NON-INVASIVE ABDOMINAL RECORDINGS ..... 197

*4.5.1 Single-Channel State-of-the-Art ..... 199*

*4.5.2 Methodology ..... 201*

*4.5.3 Experimental Results ..... 216*

*4.5.4 Discussion ..... 229*

5 CONCLUSION ..... 232

BIBLIOGRAPHY ..... 240

# List of Tables

Table 2.1 Demographic Information of Subject Groups (Average (± Standard deviation)) ..... 13

Table 2.2 Feature space by group types ..... 17

Table 2.3 Hyperparameters of the ML Models ..... 18

Table 2.4 Performance Evaluation of the Proposed Method for the Optimized and Non-optimized Models ..... 26

Table 2.5 Feature Ablation Evaluation Using XGBoost ..... 31

Table 2.6 Top-five Features of Ablation Study by SHAP Values ..... 33

Table 2.7 Comparison of the Model with Previous Works in the Literature ..... 36

Table 2.8 Demographic Information of the Healthy and VHD Participants ..... 43

Table 2.9 10-LSOCV Average (± Standard Deviation) Performance on VHD Detection for the Validation Dataset ..... 56

Table 2.10 10-LSOCV Average (± Standard Deviation) Performance on VHD Detection for the Test Dataset ..... 57

Table 2.11 Comparative Analysis of VHD Detection Based on Accelerometer Contact Microphone Data Trained on the CMT Architecture Versus State-of-the-Art ..... 65

Table 2.12 Demographic Information of the Healthy and PAD Participants ..... 69

Table 2.13 10-LSOCV Average (± Standard Deviation) Performance on PAD Detection ..... 84

Table 2.14 10-LSOCV Average (± Standard Deviation) Performance on PAD Severity Classification ..... 85

Table 2.15 Silhouette Scores of Obtained Clusters for PAD Detection Network in 10-LSOCV .. 91

Table 2.16 Silhouette Scores of Obtained Clusters for Severity Classification Network in 10-LSOCV ..... 92

Table 2.17 Comparative Analysis of PAD Detection and Severity Classification Based on Accelerometer Contact Microphone Data Trained on MSPViT Architectures Versus State-of-the-Art ..... 93

Table 2.18 Demographic Information of the Participants ..... 100

Table 2.19 Performance Evaluation for the Estimation of SBP, DBP, and MAP ..... 113

Table 2.20 The Impact of Blood Pressure Category on Estimation Performance ..... 120

Table 2.21 The Impact of Blood Pressure Level on Estimation Performance ..... 121

Table 2.22 Comparative Analysis of BP Estimation with State-of-the-Art PPG and SPG-based Methods ..... 124

Table 3.1 Performance Evaluation of the ML Techniques Used in This Work ..... 133

Table 3.2 Performance Evaluation of Abnormality Detection (0.3 < PEP/LVET < 0.4) ..... 142

Table 3.3 Estimation Accuracies Achieved for Hemodynamic Parameters ..... 154

Table 3.4 Comparative Study with Chest SCG and Wrist PPG ..... 165

Table 3.5 Performance Evaluation of the Pulse Pressure Estimation Framework. ..... 171

Table 3.6 Contributive Frequency Bands for Pulse Pressure Estimation. ..... 173

Table 4.1 RMSE of Estimation by Modality-only and Fusion-based Methods ..... 181

Table 4.2 Comparison with Other Works in Literature ..... 184

Table 4.3 Performance of the denoising method for Subject-held-out Cross-Validation for SNR= 0 $dB$ ..... 196

Table 4.4 Cross-Validation Performance for FECGSYNDB Dataset ..... 221

Table 4.5 Performance Evaluation of DPSS on the ADFECGDB and NIFECGC Datasets ..... 222

Table 4.6 Ablation Study Based on the Performance for Fetal ECG Extraction from FECGSYNDB ..... 226

Table 4.7 Comparative Evaluation for the Detection of Fetal QRS Complexes and FHR Estimation vs. State-of-the-Art Methods ..... 229

# List of Figures

Fig. 2.1 $GCG_X$ and $GCG_Y$ fiducial points and timing intervals annotated in this work..... 11

Fig. 2.2 Schematic of the experimental setup for data collection. ..... 14

Fig. 2.3 Left panel: pre-processing, right panel: feature extraction flow graphs. In the pre-processing step, motion artifacts and baseline wandering are removed from signals, which is followed by signal segmentation, peak detection, and annotation. Feature extraction is carried out for time and frequency HRV parameters as well as the GCG morphological features. Features are eventually concatenated to create a feature vector. ..... 15

Fig. 2.4 Peak detection technique based on envelope detection. ..... 16

Fig. 2.5 Beat detection from SCG and GCG channels, from top to bottom: $SCG_X$, $SCG_Y$, $SCG_Z$, $GCG_X$, $GCG_Y$, $GCG_Z$, and the envelope signal. ..... 19

Fig. 2.6 Subject-level and chunk-level datasets. In the subject-level dataset, each subject is considered a sample. In the chunk-level dataset, every chunk represents a sample. The hyperparameters of ML classifiers are fine-tuned through a 10-fold cross-validation. The trained models are evaluated using the remaining 20% of datasets. The top features are ranked according to their contribution to the classifier output. ..... 21

Fig. 2.7 Performance evaluation of XGBoost for filters based on literature. (a) Chunk-level performance for N= 10 s, (b) chunk-level performance for M= 500 ms, (c) subject-level performance for N= 10 s, and (d) subject-level performance for M= 500 ms. ..... 23

Fig. 2.8 Top features by feature importance of XGBoost and SHAP. (a) Feature importance for subject level, (b) SHAP values for subject level, (c) feature importance for chunk level, and (d)

SHAP values for chunk level. 90% and 85% consistencies are found between feature importance and SHAP values at subject-level and chunk-level, respectively. ..... 29

Fig. 2.9 The Confusion matrix for a 4-class classification. ..... 32

Fig. 2.10 Top features for the classification of aortic stenosis severity level. (a) Feature importance by XGBoost, and (b) SHAP values..... 35

Fig. 2.11 Comparison of cardiac cycle morphologies in the ACM recordings of subjects with (a) a healthy cardiac system, (b) systolic murmurs, and (c) diastolic murmurs..... 42

Fig. 2.12 The distribution of the VHD patients in terms of their coexisting valvular diseases. .... 43

Fig. 2.13 The experimental setup for the detection of coexisting valvular heart diseases. (a) The sensor head non-invasively attached to the chest wall using adhesive tape, (b) the size of the ACM sensor head compared to a dime, (c) the data acquisition module used for data transmission, (d) a two-minute ACM recording, and (e) a 10-second signal segment used to predict the sequence associated with coexisting VHDs. ..... 45

Fig. 2.14 Mel-frequency cepstral coefficients (MFCCs) and their first and second derivatives for normal cycles (a-c), systolic murmur-contaminated cycles (d-f), and diastolic murmur-contaminated cycles (g-i), respectively..... 48

Fig. 2.15 (a) the image-to-sequence translation network developed in this work for the detection of coexisting valvular heart diseases, (b) the convolutional block consisting of three convolutional layers, and (c) the convolution-transformer module (CTM)..... 51

Fig. 2.16 Receiver operating characteristic curves for the diagnosis of (a) aortic stenosis, (b) aortic regurgitation, (c) mitral regurgitation, (d) tricuspid regurgitation, (e) pulmonary regurgitation, and (f) aggregate valvular abnormalities. ..... 59

Fig. 2.17 The impact of body mass index (BMI) on the prediction scores for the detection of VHDs. ..... 63

Fig. 2.18 The proposed PAD detection framework. (a) The sensor arrangement on the torso, (b) dimensions of the accelerometer contact microphone, (c) the data acquisition module, (d) the ACM signal segmentation, and (e) decision making based on a signal segment. ..... 71

Fig. 2.19 Signal denoising examples for high-SNR and low-SNR ACM segments. (a) High-SNR ACM raw recording (SNR= 10.5 dB), (b) filtered high-SNR ACM signal segment, (c) denoised high-SNR signal segment generated by EMD, (d) low-SNR ACM raw recording (SNR= 5.3 dB), (e) filtered low-SNR ACM signal segment, and (f) denoised low-SNR signal segment generated by EMD. ..... 73

Fig. 2.20 Heart sound feature extraction. (a) denoised ACM signal segment, (b) Mel frequency cepstral coefficients (MFCCs), (c) the first derivative of MFCCs, and (d) the second derivative of MFCCs. ..... 76

Fig. 2.21 Examples of high-level-of-abstraction coefficients extracted from MFCCs and its derivatives by ResNet50. Top, middle, and bottom rows correspond to feature maps (FM1, FM10, FM20) from layers 10, 22, and 40 respectively. ..... 77

Fig. 2.22 Multi-stream-powered vision transformer (MSPViT) architecture for PAD detection network. (a) a ResNet50 for HLACs extraction followed by a vision transformer (ViT) for PAD detection, and (b) a transformer encoder which employs a multi-head attention module for modeling inter-patch dependencies. ..... 79

Fig. 2.23 Receiver operating characteristic (ROC) curves for (a) PAD binary classification, (b) healthy class in ternary classification, (c) mild-PAD class in ternary classification, and (d) severe-PAD class in ternary classification. ..... 88

Fig. 2.24 Principal component analysis (PCA) applied to the outputs of the transformer encoder corresponding to (a) MSPViT (X1), (b) MSPViT (X2), (c) MSPViT (X3), and (d) MSPViT (X, X1, X2, X3). ..... 90

Fig. 2.25 Strain plethysmography of fingertip. (a) a normal finger, (b) a dilated finger due to the blood flow, and (c) comparison between strain and optical recordings of the fingertip. ..... 99

Fig. 2.26 The proposed blood pressure estimation framework. (a) The arrangement of the fingertip strain sensor, ECG, PPG, and reference BP on the subject, (b) attachment of the strain sensor to the fingertip from two different angles, (c) the mode shape of the MEMS strain sensor, (d) the scale factor/sensitivity of the sensor is 73 fF/με, and (e) measurement protocol. ..... 101

Fig. 2.27 The visual presentation of fingertip strain recordings. (a) 2nd-order integral, (b) 1st-order integral, (c) time evolution, (d) 1st-order derivative, (e) 2nd-order derivative, (f) phase, and (g) cardiac cycle of the fingertip strain signal. ..... 104

Fig. 2.28 The distribution of systolic, diastolic, and mean blood pressure readings collected by the reference sensor from 32 subjects. ..... 110

Fig. 2.29 Scatter plots of the estimated (a) SBP, (b) DBP, and (c) MAP values in 5-CV analysis, in terms of the respective ground-truth values. ..... 114

Fig. 2.30 The Bland-Altman agreements between the estimated and reference (a) SBP, (b) DBP, and (c) MAP values in 5-CV analysis. ..... 115

Fig. 2.31 Scatter plots of the estimated (a) SBP, (b) DBP, and (c) MAP values in subject-specific analysis, in terms of the respective ground-truth values. ..... 116

Fig. 2.32 The Bland-Altman agreements between the estimated and reference (a) SBP, (b) DBP, and (c) MAP values in the subject-specific analysis. ..... 117

Fig. 2.33 The top 30 contributing features to the estimation of (a) SBP, (b) DBP, and (c) MAP. ..... 118

Fig. 2.34 Mean absolute error for the estimated SBP (left), DBP (middle), and MAP (right) values when feature categories are ablated. ..... 122

Fig. 3.1 Signal processing, feature extraction, and classification flow graph proposed in this work. ..... 128

Fig. 3.2 Fiducial points of the measured signals. From top to bottom: ECG, $GCG_X$, and $GCG_Y$. ..... 130

Fig. 3.3 Top features for AS severity classification and their average impact on the class score ranked by Light GBM. ..... 134

Fig. 3.4 The experimental setup including an accelerometer contact microphone (ACM from StethX Microsystems) on the chest connected to a data acquistion (DAQ) module, and a single lead of an electrocardiogram (ECG) sensor node (Shimmer3 ECG Development Kit) strapped around the torso. ..... 136

Fig. 3.5 The signal processing chain including pre-filtering, signal annotation, and decision-making stages for noise cancellation, time intervals estimation, and LVEF abnormality detection, respectively. ..... 138

Fig. 3.6 ACM (top) and ECG (bottom) annotations and time intervals, including pre-ejection period (PEP) and left ventricular ejection time (LVET) corresponding to time difference of Q-S1 and S1-S2, respectively. ..... 139

Fig. 3.7 The setup of wearable fingertip strain plethysmography. The strain sensor is placed on the fingertip to collect pulse information and hemodynamic parameters. (b) The fingertip pulsation signal and its fiducial points. ..... 146

Fig. 3.8 Signal segmentation and annotation of a 10-second SPG signal. (a) The band-pass filtered SPG signal. (b) the second derivative of the signal, with peaks corresponding to the inflection points in the original SPG signal. (c) The squared version of the second-derivative and the threshold for peak detection. (d) The cardiac cycles extracted from the 5-minute SPG signal..... 148

Fig. 3.9 The extraction of respiration from SPG signal. (a) A 1-minute SPG signal. (b) The SPG-derived respiration signal. (C) The reference for respiration signal from Shimmer. ..... 150

Fig. 3.10 The Bland-Altman characteristics for (a) heart rate and (b) heart rate variability, estimated by SPG and ECG. ..... 152

Fig. 3.11 The experimental setup of strain plethysmography (SPG). (a) The placement of the sensor on the fingertip and wrist for collecting pulse information. (b) The configuration of the sensitive strain sensor for data collection. (c) An example of the SPG signals recorded from the radial artery. ..... 157

Fig. 3.12 Digital and radial SPG recordings for (a) a normal subject with BMI of 20 and (b) an overweight subject with BMI of 27. ..... 159

Fig. 3.13 Mean difference and mean absolute difference for meanNN, SDNN, and RMSSD values estimated from (a) digital artery and (b) radial artery..... 161

Fig. 3.14 Mean error and mean absolute error for PNN50 and LF/HF values estimated from (a) digital artery and (b) radial artery. ..... 162

Fig. 3.15 Bland-Altman agreement between meanHR estimated from (a) digital and (b) radial arteries, with respective ECG-derived meanHR values. ..... 163

Fig. 3.16 (a) Experimental setup including an accelerometer contact microphone (ACM) attached to the pulmonary region and an expanded-coverage, preformed blood pressure cuff (OMRON HEM-FL31) secured around the left arm as reference. (b) The dimensions of the ACM sensor head. ..... 167

Fig. 3.17 Feature extraction using signal decomposition based on discrete wavelet transform (DWT). Wavelet sub-bands are used for feature extraction. ..... 168

Fig. 3.18 Detail coefficienis (DCs) and approximation coefficients (ACs) of a 10-sceond ACM signal segment achieved by 12-level DWT decomposition. ..... 170

Fig. 4.1 (a) The experimental setup for data collection: an adjustable bed, the wearable sensors, CTG machine, and band straps. (b) ECG electrodes and IMU layout on the pregnant abdomen; X, Y, and Z show the axes for IMU recordings. ..... 176

Fig. 4.2 Fusion-based FHR extraction flow graphs: (a) Early fusion, (b) Late fusion. ..... 179

Fig. 4.3 Performance of the proposed method and the sensor-alone scenarios: (a) APE, (b) PPA. ..... 183

Fig. 4.4 The generative model for discovering cross-domain relations, including two generators (***GAB*** and ***GBA*** for noisy AECG-to-denoised AECG and clean AECG-to-noisy AECG translations respectively) and two discriminators (***DA*** and ***DB*** for scoring synthetic signals generated by ***GBA*** and ***GAB*** respectively). ..... 189

Fig. 4.5 The architecture of the generators adopted for one-to-one mapping in signal-to-signal translation network. This architecture is a 1-D U-Net with three levels of max-pooling and upsampling. ..... 190

Fig. 4.6 (a) The residual discriminator network for scoring synthetic samples. (b) the residual block including three convolutional layers and a convolutional layer on the skip connection. ReLU and batch normalization (BN) are used to add non-linearity and stable convergence. ..... 191

Fig. 4.7 (a) Input noisy abdominal ECG. (b) Denoised abdominal ECG. (c) Clean abdominal ECG. ..... 194

Fig. 4.8 Root-mean-square deviation (RMSD) of the estimated signal in terms of input signal-to-noise ratio (SNR). ..... 195

Fig. 4.9 The distribution of ECG sensors on the abdominal region in FECGSYNDB dataset [255]. ..... 201

Fig. 4.10 Dual-path source separation (DPSS) framework including (a) a generative network for denoising abdominal ECG, and (b) a dual-path LSTM (DP-LSTM) network for the separation of fetal and maternal ECG signals. ..... 204

Fig. 4.11 Dual-path LSTM architecture for sequential modeling. The network includes two modeling paths, (a) intra-segment (local: path 1) and (b) inter-segment (global: path 2). ..... 207

Fig. 4.12 (a) The CGAN-based denoising network consisting of a dual-path LSTM (DP-LSTM) generator and a discriminator based upon residual blocks. (b) A residual block structure. ..... 209

Fig. 4.13 Source separation network including (a) sparse feature extraction and (b) dual-path masking blocks. The input to the network is a denoised abdominal ECG which undergoes feature

extraction and masking. The fetal and maternal ECG signals constitute the output of the network. ..... 211

Fig. 4.14 Data augmentation techniques applied on (a) abdominal ECG (AECG) from FECGSYNDB. (b) jittered AECG, (c) cropped AECG, (d) STFT of AECG, (e) masked STFT of AECG, and (f) masked ECG. ..... 213

Fig. 4.15 An 8-second test segment reconstructed by overlapping the segments of a02 from the NIFECGC dataset. (a) Noisy abdominal recording, (b) de-noised abdominal ECG, (c) extracted maternal ECG, (d) extracted fetal ECG, and (e) estimated fetal heart rate. ..... 217

Fig. 4.16 Clustering performance for FECGSYNDB, ADFECGDB, and NIFECGC datasets based on the purity, Jaccard, and Davies-Bouldin indices. ..... 223

Fig. 4.17 Impacts of (a) segment duration and (b) sub-segment duration on the F1-score obtained for the detection of fetal QRS complexes in FECGSYNDB. ..... 224

Fig. 4.18 Examples of false detection of fetal QRS, successful detection of coincided fetal and maternal QRS peaks, and missed fetal QRS. (a) abdominal signal of a16 from NIFECG, (b) de-noised abdominal signal, and (c) extracted fetal ECG signal. ..... 229

Fig. 4.19 Processing time of a 4-second abdominal ECG segment on different processors. ..... 231

# 1 Introduction

The cardiovascular system plays a critical role in maintaining optimal health and well-being, both in adults and in developing fetuses. Measuring the dynamics of blood flow within the cardiovascular system, which is referred to as hemodynamic monitoring, is essential for diagnosing and managing various cardiac conditions, including valvular heart diseases (VHD), peripheral artery disease (PAD). Similarly, continuous fetal heart rate monitoring is a vital component of antenatal care, helping to detect and manage fetal distress during pregnancy and labor.

In this chapter, we will discuss the significance of hemodynamic monitoring in the detection and management of valvular heart disease. We will explore the various methods used to monitor hemodynamics and their clinical applications, including the use of invasive and non-invasive techniques. Additionally, we will examine the pathophysiology of valvular heart disease and how hemodynamic monitoring can aid in its diagnosis and treatment.

Furthermore, we will also delve into the significance of continuous fetal heart rate monitoring in antenatal care. We will explore the various techniques used to monitor fetal heart rate and their clinical applications, including the detection and management of fetal distress. Finally, the objectives of this dissertation, which are geared towards addressing the shortcomings and limitations in the literature, will be introduced.

By the end of this chapter, readers will gain an understanding of the critical role that hemodynamic monitoring and continuous fetal heart rate monitoring play in clinical

practice. They will also gain insight into the diagnostic and therapeutic applications of these techniques in the management of valvular heart disease and fetal distress, respectively.

### 1.1 Significance of Valvular Heart Disease Detection, Peripheral Artery Disease Detection, and Hemodynamic Monitoring

Valvular heart diseases (VHD), defined as reduced or backward flow of blood through the heart valves, affect 2.5% of the US population, and incur an annual healthcare cost of $42.47 billion (2019 statistics) [1]. Early diagnosis of VHDs allows for timely therapeutic interventions such as regular screening, lifestyle modifications, and medication prescriptions to foster increased life expectancy [2]. According to a study on 919 aortic stenosis (the most common type of VHD) patients, timely interventions contributed to a significant decrease (up to 40%) in the healthcare cost of the patients [3].

Initial evaluation of patients suspected with valvular diseases is performed by a general practitioner using a stethoscope to listen to the sounds from the heart and lungs, also known as the auscultation examination. In the case that abnormal sounds (murmurs) are noticed, the patient is referred to a cardiologist who in most cases orders a transthoracic echocardiogram (TTE) test on the patient [2]. In the echocardiogram test, an ultrasound transducer non-invasively creates an image of the heart to assess the structure of the cardiac valves and heart chambers [4]. The echocardiogram is a powerful tool for the diagnosis of VHDs. According to a study performed on 213,174 aortic stenosis patients in 2019, a combination of certain TTE parameters such as the aortic valve peak velocity, mean

pressure gradient, and peak velocity ratio provided sensitivity and specificity values of up to 92% and 99%, respectively, for the detection of the valvular heart disease [5].

There are two main downsides to the evaluation procedure described above. The first relates to the low accuracy of the auscultation exam, which is subjective and susceptible to the errors of the human auditory system. As such, a significant number of VHD cases are not detected in early screening processes. In a study performed on 251 VHD patients in 2018, a low sensitivity of 44% was reported for the auscultatory exam, and this number was lower in asymptomatic patients [6]. As such, a large number of VHD cases progress to advanced stages, where outcomes are severe and treatment is costly. There is also a second downside to this process. In fall 2022, Mr. Arash Shokouhmand and his Ph.D. advisor, Dr. Negar Ebadi, participated in an I-Corps cohort to discover the needs, challenges, and desires for the diagnosis of valvular heart diseases in clinical settings. At the end of the cohort, the team realized that approximately half of the VHD patients who are referred by cardiologists for echocardiography do not need to undergo the procedure, i.e., the exam is performed out of an abundance of caution. An echocardiography exam consists of an imaging session, performed by a technician, and the interpretation of the results, performed by a cardiologist. The imaging session and the interpretation of the results take approximately 40-60 minutes and 20-45 minutes, respectively [7]. As such, a considerable amount of time and resources could be saved by alleviating unnecessary echocardiography procedures. Considering the reasons mentioned above, it becomes clear that there is an essential need for accurate and low-cost detection of valvular heart diseases at point of care.

Peripheral artery disease (PAD), defined as the narrowing or obstruction of major systemic non-coronary arteries, currently affects 8.5 million people in the US [8]. The most common cause of PAD is the deposition of plaque in the inner layer of the arteries which is either asymptomatic, or accompanied by intermittent cramping pain in the leg known as claudication [9].

Asymptomatic PAD constitutes 20%-50% of the whole PAD population, inhibiting early-stage diagnosis and risk management [10]. Hence, the disease may progress into impaired blood flow to limbs, resulting in limb soreness, infection, and gangrene, and eventually ends up with limb amputation [11]. On the other hand, early detection of PAD allows for timely intervention such as supervised exercise, medications, and revascularization to prevent adverse outcomes [12].

The ankle/brachial index (ABI) is a commonly used non-invasive test for the detection of PAD. ABI is the ratio of the systolic blood pressure at the ankle to the systolic blood pressure at the arm, suggesting the severity level of PAD [13]. A low ABI ($< 0.90$) is highly associated with the presence of occluded arteries and thus a powerful indicative of PAD [14]. Despite the ABI test being a convenient clinical practice, a reputable research conducted on 464 PAD patients in [15] reported a predictive sensitivity of 79% for the ABI test, which implies a low diagnostic performance. Angiography is the gold standard for PAD screening which offers direct visualization of the structure of an artery [16]. The procedure begins with injecting a radiographic contrast dye through a narrow, flexible catheter into the artery, which is followed by capturing X-ray images to view obstructed arterial regions with high sensitivity of 98% as reported by [17,18]. The invasive nature of

angiography however, causes local pain, puncture site, and discomfort for the patient [19]. The aforementioned shortcomings of the ABI test and angiography provoke the need for a reliable, yet convenient screening method for PAD detection.

Recent advances in wearable sensors and machine learning techniques have established the basis for the non-invasive diagnosis of PAD. Two primary methods which are currently investigated at research level include oscillometry and photo-plethysmography (PPG). In oscillometry, a blood pressure (BP) cuff is fully inflated around the subject's target limb [20,21]. The superposition of the internal arterial pressure and the external pressure by the cuff is then measured in a gradual deflation using a pressure transducer placed in the cuff. The temporal evolution of the transducer recording is analyzed by a machine learning algorithm to identify the presence of PAD. The authors in [21] trained a multilayer perceptron model on the oscillometry data collected from 14 PAD patients and 19 healthy subjects. They reported an accuracy of 91.4%, sensitivity of 90.0%, and specificity of 92.1% for the diagnosis of PAD. These values imply a missed rate of 10% and a false detection rate of 7.9% for PAD detection, questioning the prediction stability of the oscillometry method. Furthermore, the cuff used in the oscillometry technique may cause inconvenience and pain for PAD patients with claudication in the calf region.

PPG is a non-invasive method of estimating blood flow variations in arteries, where an optical sensor continuously emits light with an excitation wavelength of 680 - 950 nm to estimate the instantaneous blood volume based on the reflected energy from the artery [22]. As proven by several studies, temporal and spectral characteristics of PPG signals

recorded at toes [23–26] and fingers [24,27] can contribute to the diagnosis of PAD. According to the study conducted on 30 PAD patients in [24], it was demonstrated that the toe PPG could provide sensitivity and specificity of 0.95 and 0.65 for the diagnosis of PAD, respectively. As reported by [26], toe PPG recordings can be used with deep neural networks to achieve a sensitivity of 86.6% and specificity of 90.2% for PAD detection. However, PPG morphology can be affected by environmental lighting [23], as well as physiological characteristics such as age and health status [28,29]. A common characteristic among the aforementioned methods is the use of either a BP cuff or a PPG sensor placed on the arterial site. Both methods fail to provide acceptable prediction power as reported by [20,21,23–27].

### 1.2 Significance of Continuous Fetal Heart Rate Monitoring

The wellbeing of a fetus is generally monitored in the clinic during routine pregnancy visits. The tools commonly used are sonographic measurements of the cervical length [30], and cervical levels of fetal fibronectin [31], both with varying degrees of predictive value. Both of these require visiting a provider and are intermittent measurements over weekly or biweekly intervals. However, potential complications may well arise in between visits. Timely detection of complications is imperative, as early intervention could reduce the possibility of adverse outcomes in the prenatal period [32]. Therefore, continuous assessment of the condition of the fetus is needed outside the clinic.

Stillbirth, defined as the death of a fetus after 24 weeks of gestational age [33], affects 2.5 million babies globally every year [34]. Stillbirth is often accompanied by a decrease in fetal heart rate (FHR) [33], provoking the need for proactive fetal monitoring to reduce

fetal mortality. Current fetal heart rate monitoring predominantly relies on clinical cardiotocography (CTG) systems which are equipped with an ultrasound transducer [35–38]. However, prenatal complications such as stillbirth may arise in between clinical visits when timely medical interventions could alleviate the risk of adverse outcomes. This requires continuous tracking of fetal heart rate which serves expectant mothers to be informed of the wellbeing of the fetus. Clinical devices such as CTG are bulky and therefore not suitable for ubiquitous fetal monitoring [39].

In the past few years, the advent of wearable sensors has allowed for the dissemination of wearable health monitoring devices [40]. In parallel, a variety of technologies including electrocardiogram (ECG) [41], phonocardiogram (PCG) [42], seismo-cardiogram (SCG) [43], and gyro-cardiogram (GCG) [44] have been proposed for FHR monitoring. Non-invasive fetal ECG (NIFECG) has recently shown promise for FHR monitoring [45]. This technology incorporates ECG electrodes placed on the mother's abdomen to acquire abdominal ECG signals [46]. However, these recordings consist of several components including maternal ECG, fetal ECG, uterine contractions, fetal movements, and noise components which complicate FHR estimation [47]. Several blind source separation methods such as independent component analysis (ICA) have been proposed to extract fetal ECG from multi-channel abdominal recordings [48–50]. A necessary condition for blind source separation is that the number of channels has to be equal to or greater than the number of independent sources [51]. This implies that several ECG electrodes need to be attached to the abdomen for recording, which leads to inconvenience for pregnant women as well as large form-factor of the monitoring device.

Having analyzed the downsides mentioned above, it is crucial to utilize a wearable device to provide continuous and precise measurements of the fetal heart rate.

### 1.3 Dissertation Objectives

The primary aim of this dissertation is to investigate the practicability of creating a point-of-care apparatus that can identify valvular heart diseases, particularly aortic stenosis. Additionally, the proposed device will monitor hemodynamic parameters such as left ventricular ejection fraction and mean pressure gradient. The disseration suggests the incorporation of micro-electromechanical sensors (MEMS), specifically accelerometers and gyroscopes, to characterize the cardiovascular activities using heart-induced vibrations on the chest wall. Furthermore, acoustic and biopotential sensors will be integrated with the MEMS vibrational sensors to provide a more comprehensive evaluation of the cardiovascular system.

The secondary objective of this dissertation is to investigate the viability of using a single-channel ECG sensor and vibrational sensors positioned on the abdominal area of a pregnant woman to determine the fetal heart rate. The dissertation suggests developing advanced signal processing and pattern recognition algorithms to extract fetal ECG signal from abdominal recordings and fuse signals from multiple sources to estimate the fetal heart rate.

# 2 Diagnosis of Cardiovascular Diseases Using Recordings of Cardiac Activities on Central and Peripheral Sites

## 2.1 Diagnosis of Aortic Stenosis and Stratification of its Severity Level Using Cardiomechanical Activities on the Chest Wall

Aortic stenosis (AS), defined as the narrowing of the aortic valve opening, is among the most prevalent valvular heart diseases (VHDs) in developed countries [52]. The disease typically entails the left ventricle hypertrophy to compensate for blood outflow decrease due to stenosis [52,53]. Further decrease in the valve area results in the progressive overload of the left ventricular pressure, which eventually leads to severe AS unless medical treatment, i.e., transcatheter aortic valve replacement (TAVR) surgery [54,55], is performed. Although AS is a fairly common disease, between one-third to two-thirds of the patients remain untreated as representative symptoms such as angina remain hidden at the onset of the disease [56]. Therefore, early detection of AS requires closer monitoring of cardiac activity, wherein wearable sensors could play a crucial role [40,57–59].

In recent years, the performance of the cardiac system has been investigated through a variety of wearable technologies, including electrocardiography (ECG) [60], impedance cardiography (ICG) [61], [62], photoplethysmography (PPG) [63–66], phono-cardiography (PCG) [67], [68], ballisto-cardiography (BCG) [69,70], seismo-cardiography (SCG) [43,71,72], and gyro-cardiography (GCG) [73,74]. SCG and GCG signals, which are representatives of linear and angular vibrations of the precordium, respectively, are morphologically representative of cardio-mechanical activities [69,75,76]. They can also provide valuable information regarding cardiac timing intervals (TIs) [77–79].

SCG/GCG signals have been applied to several applications. Several works have been dedicated to cuff-less and continuous blood pressure monitoring, where the SCG signals are used for estimating the pulse arrival time or pulse transit time [80,81]. In these cases, SCG demarcates aortic valve opening (AVO) and closure (AVC) events [81]. In addition, a deep learning approach was employed to map the SCG signal into its simultaneously-recorded BCG counterpart not measurable in a wearable setting [70]. More recently, fetal heart rate (f-HR) has been extracted using SCG and GCG modalities, where promising results were achieved in comparison with concurrently-recorded fetal cardiotocography [43]. Deploying machine learning (ML) algorithms, our research group has targeted AS detection based on the SCG/GCG technology [71,72]. In these works, the time-frequency representation of all ten second segments were generated, out of which features such as the energy of frequency bands were extracted. Predictive models were then trained to classify between AS and non-AS subjects. Yet, a major drawback of the aforementioned works is the need for a reference signal, e.g., ECG [71] or PPG [72], for either signal segmentation or feature extraction. Moreover, there is no physical meaning behind the frequency-band energies which were used as features. A few other works have proposed SCG/GCG signal annotations without the need for a reference signal. For instance, some research works have proposed annotation methods based on time-domain analyses [82–84], whereas some others focus on the time-frequency characteristics of SCG/GCG signals [85].

In this research, we propose a framework for the detection of AS by employing SCG and GCG time-domain and frequency-domain morphological features in two cases, i.e.,

subject-level and chunk-level analyses. In addition, the feasibility of AS detection is investigated through heart rate variability (HRV) parameters. The proposed method does not require any auxiliary reference sensors, resulting in a convenient measurement setup. Furthermore, we introduce new time-domain features to increase the confidence level of AS diagnosis. These features are extracted through a low-complexity time-domain-based approach, where no heavy computations such as wavelet transform are needed. It is shown that the proposed features are highly correlated with the occurrence of AS. In addition, the GCG time-domain features are proven to be excellent representatives of AS, which is a promising achievement for non-invasive monitoring of the cardiac system.

### 2.1.1 Motivation

Following the occurrence of AS and in turn the changes in the forces against which the heart has to contract to eject blood, the morphology of the cardiac signals of AS patients

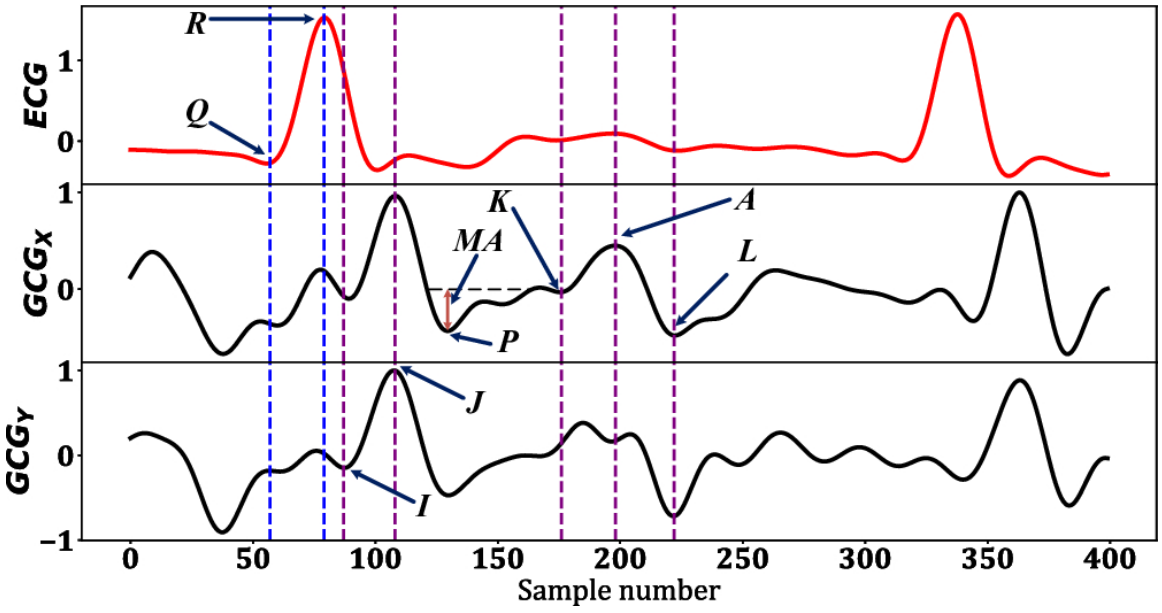


Fig. 2.1 $GCG_X$ and $GCG_Y$ fiducial points and timing intervals annotated in this work.

are expected to differ from their normal states [53]. For instance, it has been demonstrated that the progression of stenosis is conclusively correlated with ECG ST-T wave changes in a retrospective study on 29 patients [86]. Another study has recognized prolonged Q-T wave to be an independent predictor of mortality among AS patients [87]. Furthermore, the occurrence of stenosis causes HRV parameters to change accordingly [88], which will be scrutinized as a common characteristic among AS patients in this work. Fig. 2.1 depicts the ECG, $GCG_X$, and $GCG_Y$ from top to bottom, respectively, wherein $GCG_X$ and $GCG_Y$ are annotated according to the method discussed in the following sections. These axes of GCG provide useful information about the cardiac activity timing intervals as demonstrated in previous research works [71,72]. Hence, they are expected to provide potential insights into the diagnosis of AS.

### 2.1.2 Methods

#### 2.1.2.1 Experimental Protocol and Setup

This study includes thirty-two AS patients (sixteen males and sixteen females) and thirteen healthy subjects (seven females and six males). The average (standard deviation) ages of the AS and healthy groups are 84.18 (9.61) years and 68.38 (17.68) years, respectively. The demographic information of the AS and non-AS groups is summarized in Table 2.1.

Linear and angular vibrations of the chest wall were recorded using a commercial wearable sensor node (Shimmer3 from Shimmer sensing) secured by a band strap on the mid-sternum along the third rib. A three-axis accelerometer records SCG, and a three-axis gyroscope records GCG. Each modality is recorded in three dimensions: the x-axis

Table 2.1 Demographic Information of Subject Groups (Average (± Standard deviation))

| Category | Age (years) | Height (cm) | Weight (kg) |
|---|---|---|---|
| **AS** | 84.18 (± 9.61) | 164.19 (± 10.31) | 72.37 (± 13.44) |
| **Non-AS** | 68.38 (± 17.68) | 159.31 (± 35.62) | 65.94 (± 4.31) |

corresponding to the shoulder-to-shoulder direction (along the coronal axis), the y-axis corresponding to the head-to-toe direction (along the longitudinal axis), and the z-axis corresponding to the dorso-ventral direction (along the anterior-posterior axis). In this study, the dimensional letters X, Y, Z appended as sub-scripts to SCG/GCG denote the signal from the corresponding axis. Simultaneously, a four-lead ECG was recorded as a reference sensor. All waveforms were recorded synchronously at a 256 Hz sampling rate. The experimental setup is shown in Fig. 2.2.

All data were collected at the cardiac care unit of the Columbia University Medical Center (CUMC). The subjects were seated at rest on a bed or a chair for at least five minutes. They breathed naturally without controlling their breathing depths. The patient experimental protocol was approved by the Institutional Review Board of CUMC under protocol number AAAR4104. Collected data were transferred to a computer and processed in a Python framework. The flow graph of the pre-processing and feature extraction procedure is illustrated in Fig. 2.3.

#### 2.1.2.2 Signal Pre-Processing

As shown in Fig. 2.3, all channels of SCG and GCG signals were initially band-pass filtered using a 4th-order Butterworth filter with cut-off frequencies of 1-45 Hz and

1-20 Hz, respectively. Subsequently, motion artifacts associated with movements during recordings were removed using a root-mean-square (RMS) filter. In most of the literature, the RMS filter is employed by applying a M= 500 ms sliding window for signal segmentation, whereas in this work *M* is optimized as discussed in the following sections. Meanwhile, the segment removal threshold was selected twice the median value of the filter. It is worth noting that after motion artifact removal, signal chunks were not attached to each other, but processed separately. Therefore, if the duration of a chunk was less than *N* seconds, we would not take it into account for further processing. Here, *N* represents the size of the chunks, out of which the desired features are extracted.

Inspired by [84], a peak detection algorithm, as shown in Fig. 2.4, was designed to detect the GCGX and GCGY peaks and annotate the fiducial points according to Fig. 2.1. To this end, the three axes of each of SCG and GCG were combined using the root-mean-square (RMS) function, generating the linear and angular resultant vectors, respectively, as

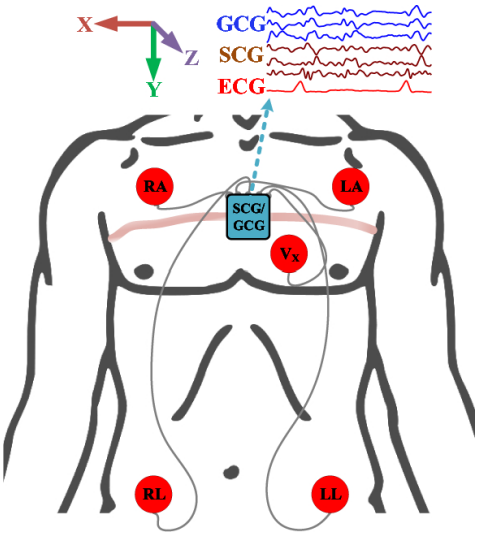


Fig. 2.2 Schematic of the experimental setup for data collection.

shown in Fig. 2.4. It is followed then by an envelope detection technique leveraging Hilbert transform [89]. Next, a 2nd-order Butterworth low-pass filter with a cut-off frequency of 2 Hz was applied to eliminate abrupt changes in the signal. Afterwards, an adaptive peak detection algorithm based on the Pan-Tompkin method was used to discriminate the real peaks in the resulted signal from summation of linear and angular envelopes [90]. Fig. 2.5 illustrates the six channels of SCG and GCG modalities followed by their corresponding beats detected by the algorithm. In the end, to locate the exact positions of the peaks in each of the 6 axes, a 50-ms window, centered at the detected peaks of the summation signal, was designed to find the local maxima associated with each axis.

After peak detection within an N-second chunk followed by a 1-10 Hz band-pass filter, signal annotation was carried out for GCGX and GCGY, where $I$, $J$, $K$, $L$, and $A$ points were characterized as proposed in [82] and marked in Fig. 2.1. Along with the

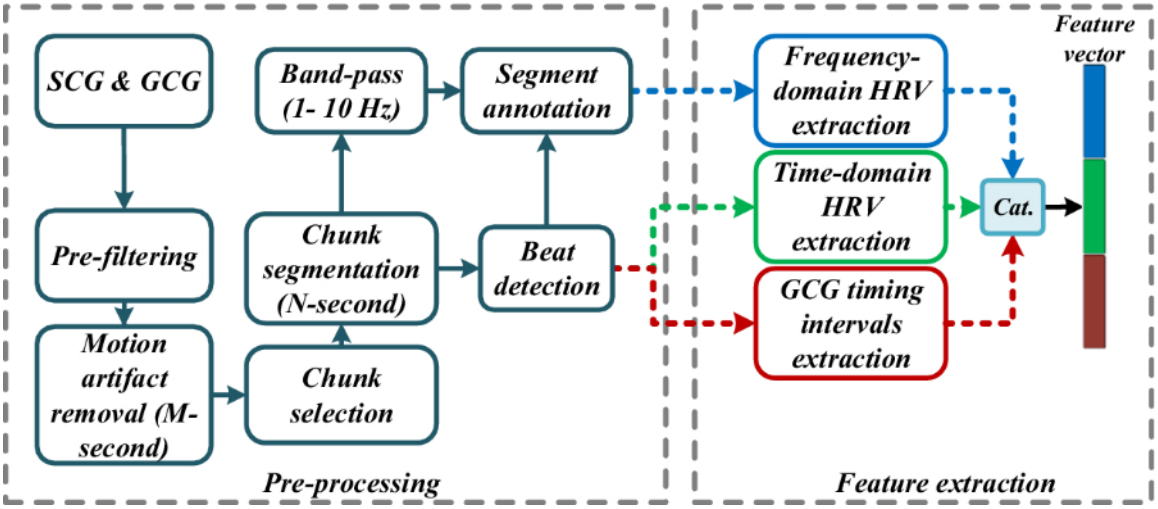


Fig. 2.3 Left panel: pre-processing, right panel: feature extraction flow graphs. In the pre-processing step, motion artifacts and baseline wandering are removed from signals, which is followed by signal segmentation, peak detection, and annotation. Feature extraction is carried out for time and frequency HRV parameters as well as the GCG morphological features. Features are eventually concatenated to create a feature vector.

mentioned points, two other parameters that we named as the maximum acceleration point (P) and its corresponding amplitude, maximum acceleration (MA), were also extracted. According to the reasons mentioned earlier, we hypothesize that the time-domain features could correlate with the anatomical changes caused by the occurrence of AS.

#### 2.1.2.3 Feature Extraction

Three types of features were extracted from SCG and GCG signals: time-domain HRV parameters, frequency-domain HRV parameters, and GCG timing interval parameters.

**Time-domain HRV parameters:** Time-domain HRV parameters provide valuable information about the cardiac activity [91]. For temporal HRV parameters, a few time-domain analyses were applied to the series of successive inter-beat intervals (IBIs). The normal-to-normal IBI (NN) is defined as the interval between consecutive *J* peaks in the GCG signal [92]. A few HRV parameters were extracted from the NN time series, such as the average (AVNN), the standard deviation (SDNN), root-mean-square of successive differences (RMSSD), and the proportion of the number of adjacent NN intervals whose durations differ more than 50 ms (NN50) to the total number of NNs (pNN50). It is worth

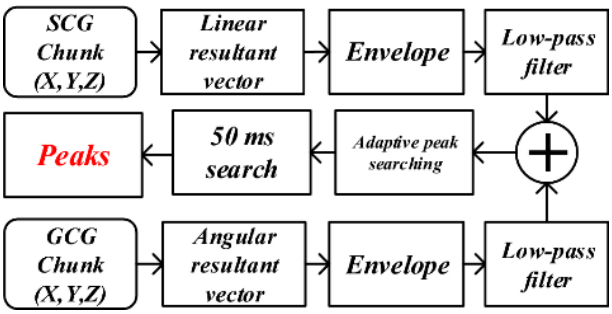


Fig. 2.4 Peak detection technique based on envelope detection.

mentioning that SDNN, RMSSD, and pNN50 are of great clinical significance as they allow for measuring cardiac risk, respiratory arrhythmia, and parasympathetic nervous activity [91,93]. Additionally, to further explore the impact of NN on AS detection, we introduced the median, skewness, kurtosis, entropy (ENN), self-entropy (SENN), and conditional entropy (CENN) values associated with NNs as new features. Due to the nonlinearity underlying the dynamics of HRV, we also extracted the vector angular index (VAI), the vector length index (VLI), SD1, and SD2 out of Poincare map - a scatter plot of NN at time t in terms of NN at time t+1 [94]. These features were calculated according to (1) to (4):

Table 2.2 Feature space by group types

| **Time HRV features** | **GCG features*** | **Frequency HRV features** |
|---|---|---|
| AVNN | IVCT (J – I) | VLF |
| SDNN | LVET (K – J) | LF |
| RMSSD | IVRT (L – K) | HF |
| pNN50 | L – I | LF/HF |
| SD1 | L – J | Total power |
| SD2 | K – I | |
| SD1/SD2 | P – I | |
| VAI | P – J | |
| VLI | K – P | |
| ENN | L – P | |
| CENN | MA | |
| SENN | | |
| NN skewness | | |
| NN kurtosis | | |
| NN median | | |

* Mean, median, standard deviation, skewness, kurtosis, entropy, min, and max were calculated for every parameter in this column.

$$\mathrm{VAI} = \frac{1}{B}\sum_{k=0}^{B}|\theta_i - 45°|, \qquad (1)$$

$$\mathrm{VLI} = \sqrt{\frac{1}{B}\sum_{k=0}^{B}|l_i - L|^2}, \qquad (2)$$

$$\mathrm{SD1} = \mathrm{std}\left(\frac{|NN_{i+1} - NN_i|}{\sqrt{2}}\right), \qquad (3)$$

$$\mathrm{SD2} = \mathrm{std}\left(\frac{|NN_{i+1} - NN_i|}{\sqrt{2}} - 2 \times AVNN\right), \qquad (4)$$

where $\theta_i$ denotes the angle of the $i_{th}$ scatter point with the x-axis. $l_i$ and $L$ indicate the distance of the $i_{th}$ point to the origin and mean of distances of all $B$ points to the origin, respectively.

**Frequency-domain HRV parameters:** Frequency-domain analysis was carried out based on the estimation of power spectral density (PSD) of the SCG and GCG signals, where the oscillation power of very-low-frequency (VLF), low-frequency (LF), and high-frequency

Table 2.3 Hyperparameters of the ML Models

| DT | RF | XGBoost | SVM |
|---|---|---|---|
| maximum depth | maximum depth | maximum depth | regularization parameter |
| minimum samples per leaf | minimum samples per leaf | minimum samples per leaf | kernel |
| minimum split samples | minimum split samples | minimum split samples | kernel coefficient |
| maximum features | maximum features | maximum features | class weights |
| criterion | criterion | criterion | |
| class weights | number of estimators | learning rate | |
| | class weights | class weights | |

(HF) bands were extracted as frequency-domain HRV parameters. It was shown that parasympathetic activities are manifested in HF (0.15-0.4 Hz), whereas sympathetic activities belong to the LF (0.04-0.15 Hz) as well as VLF (0.0033-0.04 Hz) ranges [95]. In addition to the mentioned features, the total power of PSD was calculated as an additional feature.

**GCG timing intervals:** A few timing intervals describing the cardiac system were calculated for the GCG signal. It was demonstrated that the isovolumetric contraction time (IVCT), isovolumetric relaxation time (IVRT), and left ventricular ejection time (LVET) could be estimated using *J-I*, *L-K*, and *K-J* [82]. Furthermore, we investigated whether

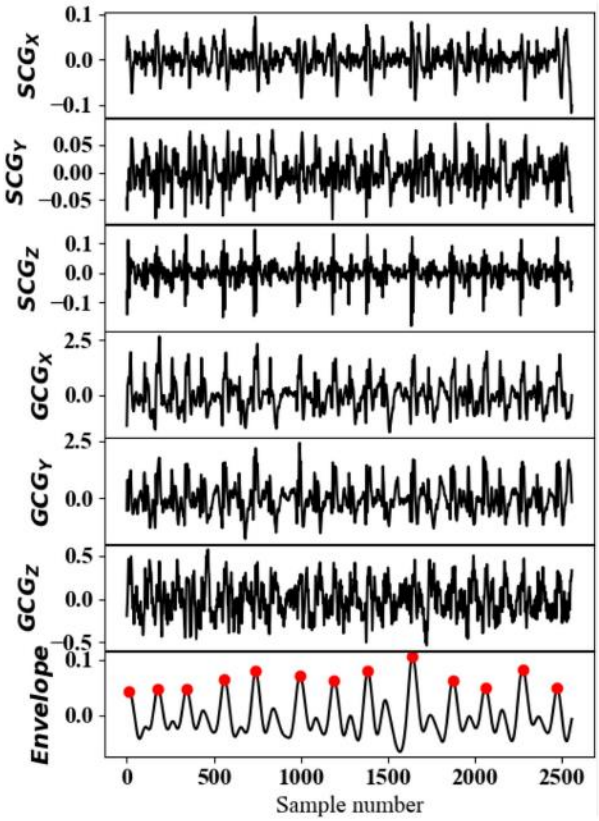


Fig. 2.5 Beat detection from SCG and GCG channels, from top to bottom: $SCG_X$, $SCG_Y$, $SCG_Z$, $GCG_X$, $GCG_Y$, $GCG_Z$, and the envelope signal.

MA, P, and its corresponding intervals have any impact on AS detection. Other parameters such as the intervals between each pair of the fiducial points depicted in Fig. 2.1 along with their mean, median, standard deviation, skewness, and kurtosis values were extracted as auxiliary features. The logic behind such an exhaustive feature extraction is to characterize the most relevant GCG timing intervals resulting in the highest accuracy of AS diagnosis. The extracted features are summarized in Table 2.2.

#### 2.1.2.4 Training and Hyperparameter Optimization

Fig. 2.6 shows the general schematic of the proposed machine learning framework including feature engineering, data split, training, and feature selection. As demonstrated in this figure, AS detection was performed in two different cases: subject-level and chunk-level. The former implies using each subject as a single sample, whereas the latter suggests each signal chunk as a sample for training the predictive models. In the chunk-level feature space, frequency-domain features were avoided since the chunk size is not long enough to accurately calculate the spectral parameters, while the whole signal could be used to measure the spectral features. Regardless of the scenario, the entire feature space was split into two parts, training (80%) and test (20%) datasets. Following the data split, we trained the predictive models, where the hyperparameters were optimized through leave-one-out 10-fold cross-validation (10-CV). For subject-level 10-CV, 0.1 of the subjects were held out at each fold, and the model continued to be optimized using the rest 0.9 held-in subjects. However, in the chunk-level 10-CV training, 0.1 of the total chunks were held out. It should be noted that both training and hyperparameter optimization were carried out using only the training part in each case.

### 2.1.2.5 Classification Techniques and Evaluation Metrics

Given the two datasets, we used four machine learning techniques for the diagnosis of AS: decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), and support vector machine (SVM) [96,97]. These classifiers are widely employed in a variety of biomedical applications including chronic kidney disease (CKD) diagnosis [98], heart failure prediction [99], lung sound auscultation [100,101], and cancer prognosis [102] due to their robustness.

During the 10-CV, the hyperparameters of each model were tuned in terms of performance. Table 2.3 presents the parameters with respect to which the ML models were optimized. For instance, the DT model was optimized in terms of the maximum depth of the tree, the minimum number of samples in a leaf, the minimum number of samples for splitting a node, the maximum number of features, and the criterion used for root selection.

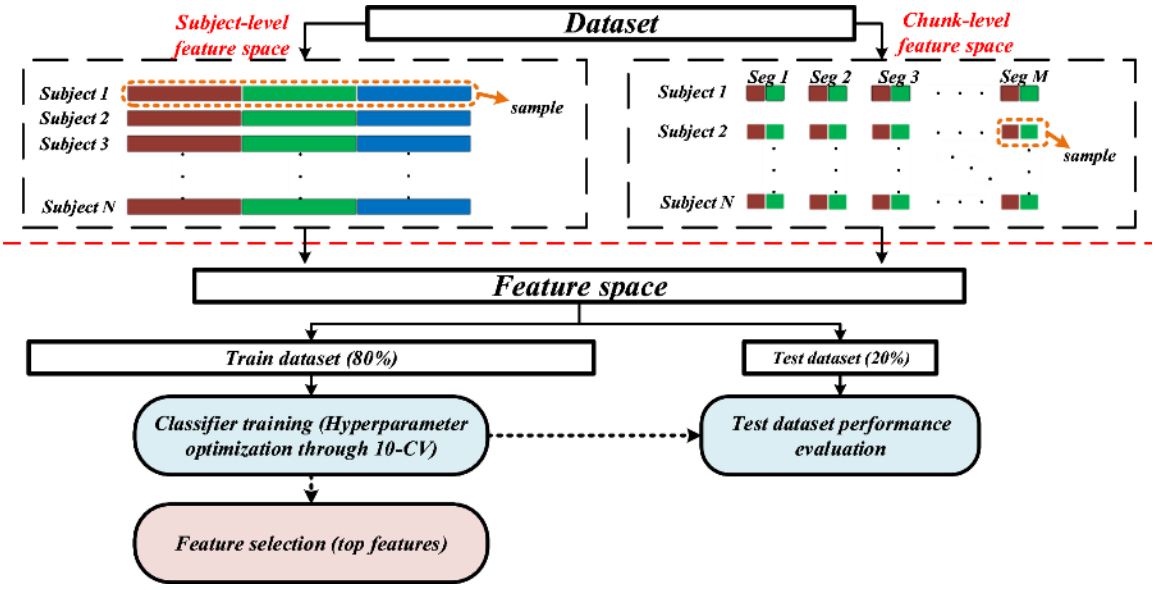


Fig. 2.6 Subject-level and chunk-level datasets. In the subject-level dataset, each subject is considered a sample. In the chunk-level dataset, every chunk represents a sample. The hyperparameters of ML classifiers are fine-tuned through a 10-fold cross-validation. The trained models are evaluated using the remaining 20% of datasets. The top features are ranked according to their contribution to the classifier output.

Similar parameters were optimized for RF and XGBoost. For the case of XGBoost, however, the model should also be tuned in terms of the learning rate, since training the XGBoost follows a gradient-based pattern. Furthermore, all predictive models were optimized in terms of class weights to tackle the data imbalance.

The following metrics were calculated to evaluate the performance of the classification algorithms: precision ($PR$), recall ($RE$), accuracy ($AC$), and F1-score ($F1$).

$AC$ is a simple metric that measures the accuracy of the model in prediction, and deals only with true predictions. However, it does not provide information for the cases where the model misclassifies a sample. $PR$ and $RE$ serve to deal with this problem by introducing false alarms and missed rates, respectively. As a combination of $PR$ and $RE$, $F1$ offers a more comprehensive understanding of the model performance, and was used for filter optimization as well.

#### 2.1.2.6 Filter Optimization

As explained in the previous sections, we employed an RMS filter for motion artifact removal with length *M* (ms). It was also mentioned that we segmented the GCG signals into N-second windows. These parameters tend to be fixed to M= 500 ms and N= 10 s in the literature [72,103], which are experimental numbers. We analyzed the effects of changing these parameters on the performance of XGBoost model. To this end, the performance metrics were assessed in terms of different values for *M* and *N* in Fig. 2.7, separately. Fig. 2.7 (a) and (b) illustrate the performance for N= 10 s in terms of different values for *M* and M= 500 ms in terms of different values for *N* at the chunk-level analyses,

respectively. The same characteristics are investigated in Figs 2.7 (c) and (d) for the subject-level. As depicted in these figures, suggesting optimum values for *M* and *N* is not a straightforward process. For instance, Fig. 2.7 (a) suggests N= 10 s and M= 500 ms for the maximum F1-score, whereas in Fig. 2.7 (b) in order to achieve the best performance, *N* should hold 18 s, which in turn implies that these parameters need to be tuned at the same time. We, therefore, defined an optimization problem to simultaneously optimize the values of *M* and *N*. The optimization problem is defined as follows:

$$\begin{cases} M^*, N^* = \underset{M,N}{argmax}\{F_1(M,N)\} \\ s.t.\ \ 100\ ms \leq M \leq 2000\ ms, \\ 2\ s \leq N \leq 25\ s \end{cases} \qquad (5)$$

where *M** and *N** indicate the optimum values resulting in the highest possible F1-score. To solve the optimization problem, the Bayesian optimization method is employed. This optimization method introduces a surrogate for the cost function and measures its

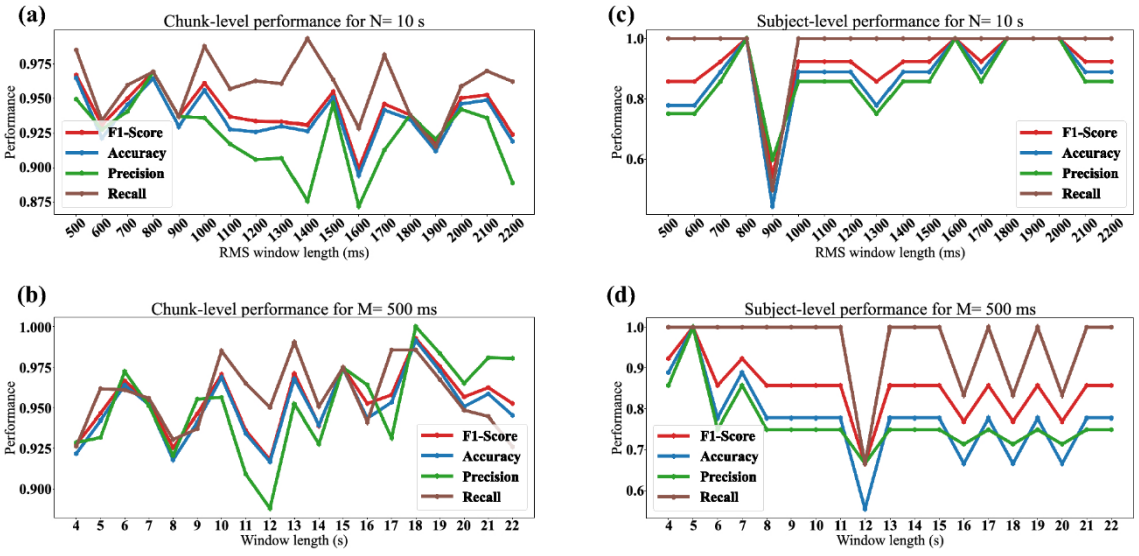


Fig. 2.7 Performance evaluation of XGBoost for filters based on literature. (a) Chunk-level performance for N= 10 s, (b) chunk-level performance for M= 500 ms, (c) subject-level performance for N= 10 s, and (d) subject-level performance for M= 500 ms.

uncertainty by a Bayesian learning technique and a Gaussian process regression. It then defines an acquisition function from the surrogate to determine the sampling locations and find the optimum values [104]. In this way, an end-to-end training procedure results in the optimized objective function and its corresponding filter parameters.

### 2.1.3 Experimental Results and Discussion

In this section, the experimental results are discussed and compared with the literature.

#### 2.1.3.1 Datasets, Features, and Filter Optimization

As mentioned earlier, two datasets were provided. The subject-level dataset included 45 subjects, where 36 of them were used to train the models, and the remaining 9 were held out for the test datasets. The held-out group consists of two mild AS, one moderate AS, two severe AS, and four healthy individuals. Out of 36 training subjects, at each fold of 10-CV, 32 subjects were held in for training and 4 subjects were held out for hyperparameter optimization. Per each subject, mean, median, standard deviation, skewness, kurtosis, entropy, min, and max (8 features) were calculated for every 11 GCG features tabulated in Table 2.2. Thus, the total number of GCG features was $11\times8 = 88$. Furthermore, 5 frequency-domain HRV parameters were extracted for each channel axis (6 axes) per subject, amounting to $6\times5=30$ features. Besides, 15 time-domain HRV parameters were extracted out of NNs for each subject. Therefore, the total number of extracted features for each sample in the subject-level feature space was 123. Chunk-level feature space included 1,272 chunks, out of which 88 GCG timing parameters as well as 15 time-domain HRVs (total of 103 features) were calculated. For training the chunk-level

dataset, 255 and 1017 chunks were used for test and training, respectively. Subsequently, 102 chunks were held out for 10-CV hyperparameter optimization, whereas 915 chunks were held in for fine-tuning the model.

The RMS filter length and chunk window length were optimized by the Bayesian optimization technique through the process of training. It should be noted that these parameters were optimized by involving the training dataset, but not the test dataset. The optimum values achieved for filter parameters were $M^*$=1582 ms and $N^*$=11.2 s. The RMS filter length is three times the value proposed in the literature, i.e., $M$=500 ms, whereas the length of the chunk is almost consistent with the literature.

### 2.1.3.2 Feature Selection for Models

We conducted both the Shapley Additive exPlanations (SHAP) technique, a game theory-based approach for interpreting the output of a model in terms of the input feature space, and the feature-importance f-score from XGBoost to provide an explainable model [105]. SHAP is employed as it calculates the feature importance in terms of the impact of every single observation on the output performance, whereas the f-score from XGBoost represents the importance of a feature for the whole dataset. The score for each feature in

Table 2.4 Performance Evaluation of the Proposed Method for the Optimized and Non-optimized Models

| **Classification level** | **Subject (Optimized parameters)** | | | | **Chunk (Optimized parameters)** | | | |
|---|---|---|---|---|---|---|---|---|
| **Classifier** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
| **DT** | 100.00 | 100.00 | 100.00 | 100.00 | 87.94 | 90.51 | 87.32 | 88.89 |
| **RF** | 66.66 | 66.66 | 100.00 | 80.00 | 88.21 | 87.88 | 92.95 | 90.34 |
| **XGBoost** | **100.00** | **100.00** | **100.00** | **100.00** | **96.36** | **95.65** | **97.35** | **96.49** |
| **SVM** | 100.00 | 100.00 | 100.00 | 100.00 | 88.72 | 87.42 | 92.96 | 90.09 |
| **Classification level** | **Subject (Literature-based parameters)** | | | | **Chunk (Literature-based parameters)** | | | |
| **Classifier** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
| **DT** | 77.78 | 87.50 | 87.50 | 87.50 | 81.96 | 84.14 | 84.14 | 84.14 |
| **RF** | 66.66 | 66.66 | 100.00 | 80.00 | 84.71 | 84.87 | 88.97 | 86.87 |
| **XGBoost** | **77.78** | **75.00** | **100.00** | **85.71** | **94.89** | **95.83** | **95.17** | **95.49** |
| **SVM** | 77.78 | 71.43 | 100.00 | 83.33 | 85.09 | 87.41 | 86.21 | 86.81 |

SHAP method is calculated by the following formula:

$$\alpha_i = \sum_{\gamma \subset \Gamma \setminus i} \frac{|\gamma|!(|\Gamma|-|\gamma|-1)!}{|\Gamma|!} [F1(\gamma \cup i) - F1(\gamma)], \qquad (6)$$

where $\alpha_i$, $\gamma$, $i$, and Γ denote the score value of the $i^{th}$ feature, any subset of feature space excluding the $i^{th}$ feature, $i^{th}$feature, and the entire feature space, respectively. $F1$ also represents the F1-score, where $[F1(\gamma \cup i) - F1(\gamma)]$ indicates the difference of F1-score resulting from incorporating the $i^{th}$ feature within the feature space. For every feature, $\alpha_i$ is calculated for all samples in the training dataset and the resulting values are averaged to obtain the feature score. F-score by XGBoost, on the other hand, evaluates each feature based on its impact on the output for all samples at once. Thus, SHAP values validate the feature ranking obtained by XGBoost feature importance and vice versa. That is, if the scores from both cases demonstrated an agreement, one can conclude the model is robust enough for the prediction task. In the following sections, we assess the importance of the features through both XGBoost and SHAP values.

#### 2.1.3.3 Performance Evaluation

After training and hyperparameter optimization, the performance of the proposed method is evaluated by applying the model on the test dataset to predict the existence of AS. Next, by comparing the predicted values and the true labels, the performance of the method is reported using metrics introduced in the previous section. Table 2.4 summarizes the performance results for the two datasets. The optimized method was reported by 100% accuracy, 100% F1-score, 100.00% precision, and 100.00% recall for DT, XGBoost, and SVM at the subject-level. As reported, DT outperforms RF, indicating 77.78% accuracy and 87.50% F1-score for the models with literature-based parameters ($M$=500 ms and $N$= 10 s). Similarly, DT suggests a higher F1-score in comparison to RF, i.e., 100.00% vs. 80.00%, for the models with optimized parameters ($M^*$= 1582 ms and $N^*$= 11.2 s).

Having a closer look at the performance of the models with optimized parameters and literature-based ones at subject-level analyses, the accuracy and F1-score of DT were improved, respectively, by 22.22% and 12.50%. Furthermore, XGBoost shows 85.71% F1-score, which is the highest performance among the models with literature-based parameters at the subject level. Yet, XGBoost with the optimized filter parameters implies 100.00% of accuracy and F1-score at the subject level. Such an improvement demonstrates the functionality of filter optimization. According to the optimized chunk-level analyses in Table 2.4, XGBoost and RF outperform other methods, suggesting 96.49% and 90.34% of F1-score, respectively. These values offer a more robust performance than the non-optimized F1-score, i.e., 95.49% for XGBoost and 86.87% for RF. Furthermore, the F1-score of the optimized XGBoost model degrades from 96.49% to 95.49% when it uses the literature parameters, denoting an increase in FP as confirmed by the decrease in recall (97.35% vs. 95.17%). In the chunk-level dataset, XGBoost surpassed the other methods at high margins, i.e., 7.64%, 5.14%, 4.391%, and 6.15% for accuracy, precision, recall, and F1-score, respectively. RF performs better than DT in terms of F1-score (90.34% vs. 88.89%), suggesting that the larger the dataset, the easier the generalizability of the RF model to the test datasets. Nevertheless, XGBoost offers a superior predictive characteristic than other methods at the chunk level. Additionally, the other optimized models at chunk-level analyses outperform their non-optimized versions in terms of F1-score by an acceptable margin (DT: 88.89% vs. 84.14%, RF: 90.34% vs. 86.87%, and SVM: 90.09% vs. 86.81%).

Fig. 2.8 depicts the top 20 features for both chunk and subject levels. Figs 2.8 (a) and (b) illustrate the top 20 features achieved by XGBoost f-score and SHAP values for subject-level analyses, respectively. Fig. 2.8 (c) and d show the same order for chunk-level analyses. Among the top 20 features of subject-level analyses, eighteen features (90.00%) are consistent between the two criteria of importance, whereas seventeen features (85.00%) are reported as the common ones between the two methods at the subject level. According

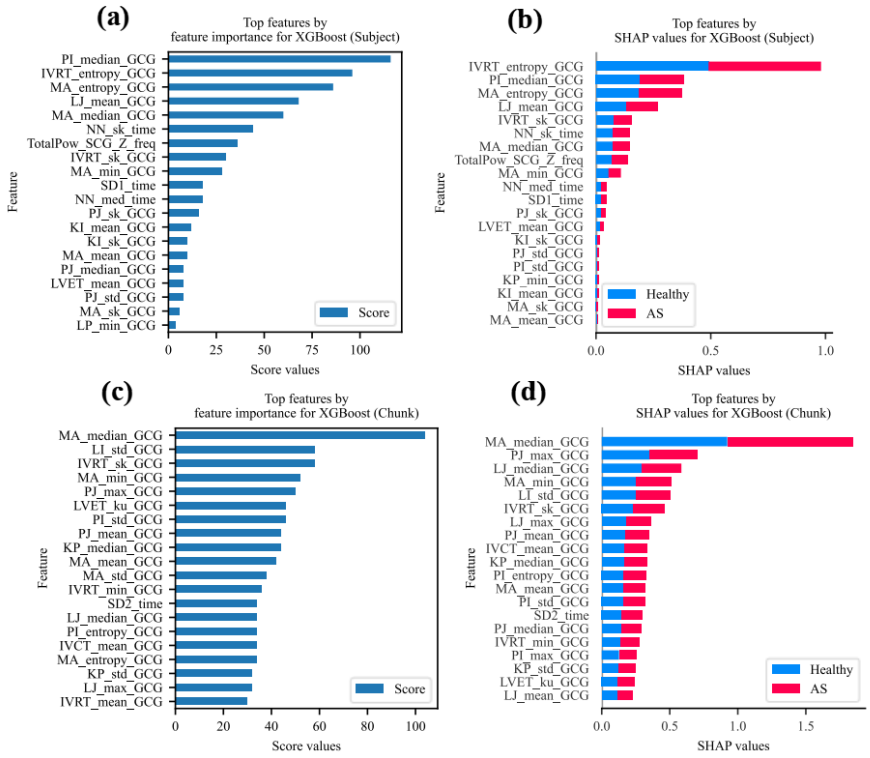


Fig. 2.8 Top features by feature importance of XGBoost and SHAP. (a) Feature importance for subject level, (b) SHAP values for subject level, (c) feature importance for chunk level, and (d) SHAP values for chunk level. 90% and 85% consistencies are found between feature importance and SHAP values at subject-level and chunk-level, respectively.

to Fig. 2.2 (b), among the top features, sixteen belong to GCG morphological features, three to time-domain HRV, and one to frequency-domain HRV parameters, demonstrating the higher representativeness of GCG features in comparison to others. Moreover, the

highest occurrences belong to MA and P-point-included intervals, each with five features among the top-20's. From the time-domain HRV parameters, NN shows the highest correlation with AS, whereas the only feature is the total power of $SCG_Z$ from the frequency-domain HRV space. For the chunk-level analyses, nineteen features belong to the GCG time-domain features, whereas only one feature, i.e., NN, belongs to the time-domain HRV feature space. Among the top GCG features, MA, P-J, and L-J repeat three times, suggesting the highest frequencies among the top 20 features. Furthermore, the top features presented in Figs 2.8 (b) and (d) almost equally contribute to making a decision on AS and healthy states (blue and red bars), denoting that the top features are representative of both classes.

#### 2.1.3.4 Ablation Study on Feature Space

An ablation study was conducted on feature spaces to investigate the predictive performance of AS diagnosis. XGBoost was used in this section due to its superior performance on both datasets. The optimized parameters were used for both subject-level and chunk-level classification tasks. Three feature spaces for the subject-level and two feature spaces for the chunk-level (frequency-domain HRV parameters are not included for chunk-level) datasets were assessed. Thus, the performance of the model was evaluated for GCG time-domain parameters, time-domain HRV, and frequency-domain HRV parameters separately. The results are summarized in Table 2.5. As shown for the chunk-

Table 2.5 Feature Ablation Evaluation Using XGBoost

| **Level** | **Subject** | | | | **Chunk** | | | |
|---|---|---|---|---|---|---|---|---|
| **Feature type** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
| **GCG intervals** | **100.00** | **100.00** | **100.00** | **100.00** | **95.00** | **93.97** | **96.46** | **95.19** |
| **Time HRV** | 55.56 | 75.00 | 50.00 | 60.00 | 73.64 | 73.91 | 75.22 | 74.56 |
| **Frequency HRV** | 66.67 | 66.67 | 100.00 | 80.00 | N/A | N/A | N/A | N/A |

level study, the best performance belongs to GCG timing intervals (95.19% F1-score), whereas the weakest performance is achieved by time-domain HRV parameters (74.56% F1-score). Furthermore, frequency-domain HRV introduces more relevant information for AS detection compared to time-domain HRV features at the subject-level (80.00% vs. 60.00%). This implies that frequency-domain HRV parameters carry more information regarding AS than the time-domain HRV parameters (accuracy: 66.67% vs. 55.56%).

Furthermore, the top five features for the datasets in the ablation study are extracted and shown in Table 2.6. According to the time-domain HRV parameters listed for chunk-level and subject-level studies, NN, SD1, and SD1/SD2 contribute more than others to AS diagnosis, whereas the top features from frequency-domain HRV features include the total power, LF/HF ratio, and HF power. Interestingly, three GCG features out of five in both cases are associated with MA and the intervals involving the P-point. Also, IVRT is ranked twice as the best feature for the subject level, showing the importance of this feature.

### 2.1.3.5 Classification of Severity Level

In this section, the robustness of the proposed method is evaluated for diagnosing the severity level of aortic stenosis on a per-patient basis. For this purpose, a 4-class classification was conducted to observe the failure modes of the proposed method. The classification involves the healthy group as well as mild, moderate, and severe aortic stenosis cases. For a 4-class classification, we augmented the size of the dataset by considering 50% of overlap between every two consecutive signal chunks. This practice helps the predictive model to better be generalized to the test dataset. As a result, a total of 2,336 chunks were generated, of which 1,868 and 468 were used for training and test datasets, respectively.

The 4-class classification results suggest 92.72% precision, 91.95% recall, 93.80% accuracy, and 92.29% F1-score. As for the F1-score, the performance of the 4-class classification has been slightly degraded compared to the binary classification for aortic stenosis (92.29% vs. 96.49%). This small difference has occurred due to the small size of

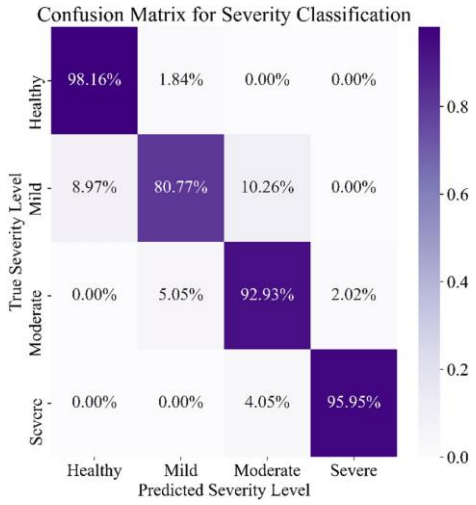


Fig. 2.9 The Confusion matrix for a 4-class classification.

the dataset with respect to the number of classes (4 classes). To investigate the failure modes, the classification results are summarized in a confusion matrix depicted in Fig. 2.9. As illustrated in this figure, 98.16% of the healthy group were classified correctly, whereas the remaining 1.84% were classified as mild cases. According to the confusion matrix, the mild, moderate, and severe cases were classified correctly with the rates of 80.77%, 92.93%, and 95.95%, respectively. These results suggest that severe patients' recordings hold distinctive characteristics which serve to distinguish them from other cases. The 4.05% misclassified severe cases were reported as moderate cases. This misclassification has happened due to the morphological similarities between moderate and severe cases. Interestingly, a severe case was never misclassified as a healthy case or mild AS, which demonstrates the robustness of the proposed feature space with respect to the number of classes. As for the mild severity, 8.97% and 10.26% were classified as healthy and moderate cases, respectively, which could have been caused by either the size of the training dataset, or the similar morphological characteristics of mild AS with healthy and

Table 2.6 Top-five Features of Ablation Study by SHAP Values

| **Subject-Level** | | | **Chunk-Level** | | |
|---|---|---|---|---|---|
| **Time HRV features** | **GCG features** | **Frequency HRV** | **Time HRV features** | **GCG features** | **Frequency HRV** |
| NN_sk_time | IVRT_entropy_GCG | TotalPow_GCG_X_freq | NN_med_time | MA_median_GCG | N/A |
| NN_ku_time | IVRT_sk_GCG | lfhfRatio_GCG_Z_freq | AVNN_time | LJ_median_GCG | N/A |
| SD2_time | PI_median_GCG | hf_HRV_GCG_Y_freq | SDNN_time | PJ_max_GCG | N/A |
| SD1/SD2_time | MA_min_GCG | TotalPow_GCG_Z_freq | SD1/SD2_time | MA_mean_GCG | N/A |
| AVNN_time | MA_entropy_GCG | hf_HRV_GCG_X_freq | SD1_time | LJ_max_GCG | N/A |

moderate cases. As for the moderate cases, a small aggregate of 7.07% was misclassified as either mild or severe, which proves the practicality of the proposed framework. Having considered the aforementioned points, it is concluded that the proposed method offers high reliability in detecting and classifying AS.

Similar to AS detection task, the top features for the 4-class classification were obtained through the feature importance and the SHAP value methods, as summarized in Figs 2.10 (a) and (b), respectively. By comparing the two feature sets, a high consistency can be observed in 16 common features, most of which are from the GCG timing intervals. Fig. 2.10 (b) also contributes to interpreting the impact of each feature on the predicted classes. As shown, 5 MA-related features are among the top features, which demonstrates their importance in classifying severity level. According to Fig. 2.10 (b), MA_mean, IVRT_median, SDNN, and PJ_skewness indicate higher consistencies with severe cases than other features, which can be considered for future studies on AS. Certain features such

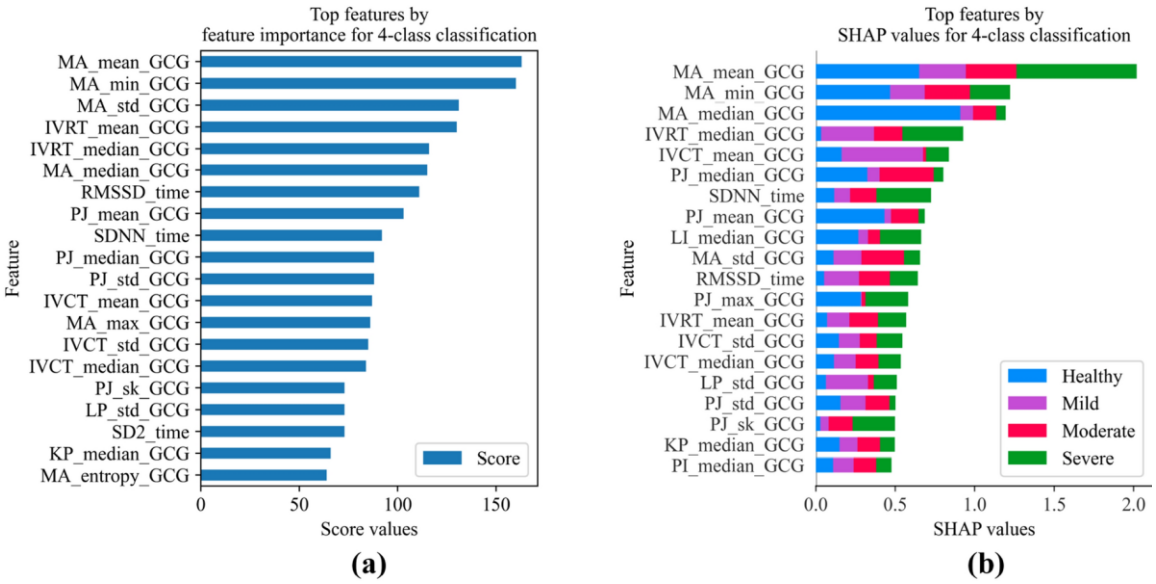


Fig. 2.10 Top features for the classification of aortic stenosis severity level. (a) Feature importance by XGBoost, and (b) SHAP values.

as SDNN, MA_std, IVCT_std, and IVCT_median contribute to both healthy and mild classes equally, which may explain the 8.97% of the mild cases misclassified as healthy.

### 2.1.3.6 Comparison with Literature

AS detection is considered only in a limited number of research works, and this motivates the authors to compare the proposed framework with the works addressing heart disease detection/classification through SCG/GCG modalities. Therefore, the proposed method is compared with other methods in the literature for detecting AS [71,72], as well as the works using SCG/GCG for the detection of other types of cardiovascular diseases (CVDs) including atrial fibrillation (AFib) and acute myocardial infarction [103,106–108]. The results are summarized in Table 2.7. The methods are compared in terms of their target CVDs, classifier, performance metrics, need for auxiliary signals other than SCG/GCG for feature extraction, and computational complexity. It should be noted that the proposed

Table 2.7 Comparison of the Model with Previous Works in the Literature

| Signal type | Target CVD | Classifier | Accuracy | Precision | Recall | F1-score | Auxiliary | Computation | Reference |
|---|---|---|---|---|---|---|---|---|---|
| **SCG + GCG** | Acute myocardial | Kernel SVM | N/A | 0.95 | 0.82 | 0.88 | No | Low | [106] |
| **SCG + GCG** | AFib | Kernel SVM | 0.98 | N/A | 0.97 | N/A | No | Low | [103] |
| **SCG + GCG** | AFib | Kernel SVM | 0.95 | 0.92 | 0.90 | 0.96 | No | Low | [107] |
| **SCG + GCG** | AFib | Kernel SVM | 0.97 | 1.00 | 0.93 | 0.96 | No | Low | [108] |
| **SCG + GCG** | AS | RF | 0.98 | 0.99 | 0.98 | 0.98 | Yes (ECG) | High | [71] |
| **SCG + GCG** | AS | XGBoost | 0.93 | 0.95 | 0.95 | 0.95 | No | High | [72] |
| **SCG + GCG** | **AS** | **XGBoost** | **1.00** | **1.00** | **1.00** | **1.00** | **N o** | **Low** | **Proposed (subject)** |
| **SCG + GCG** | **AS** | **XGBoost** | **0.96** | **0.95** | **0.97** | **0.96** | **N o** | **Low** | **Proposed (chunk)** |

method is of low computational complexity due to using merely single-dimensional signal analyses, whereas some of literature methods which offer high performances tend to employ two-dimensional transformations, such as continuous wavelet transform (CWT) [71,72]. The results of our method at the subject level are superior to those of all previous works (F1-score: 1.00 vs. 0.88 [106], 0.95 [72], 0.96 [107], 0.96 [108], and 0.98 [71]), despite the low computation power required by our method. It stands to reason that the lower the complexity, the more applicable the algorithm to real-time applications, which is a crucial point for wearable sensing systems. Even at the chunk level, our method performs equally or more accurately than other methods (F1-score: 0.96 vs0.88 [106], 0.95 [72], 0.96 [107], and 0.96 [108]), except for the method proposed in [71]. However, our method predicts AS without any auxiliary signals, whereas in [71], the ECG signal was used for feature extraction. Furthermore, the proposed framework suggests an analogous or higher

F1-score, i.e., 0.96, and recall, i.e., 0.97, in comparison to the methods for detection of other CVDs. Furthermore, the proposed framework goes further beyond the disease detection by offering explainable modeling using SHAP values and feature importance from XGBoost to extract the physical meaning of the readings. To recapitulate, one could appreciate the proposed method as an accurate, computationally-efficient, and interpretable approach for AS detection which contributes to low-cost wearable sensing systems.

## 2.2 Diagnosis of Coexisting Valvular Heart Diseases Using Image-to-Sequence Translation of Contact Microphone Recordings

Valvular heart disease (VHD), defined as reduced or backward flow of blood through heart valves, affects 2.5% of the US population [1], involving an annual cost of $42.47 billion for healthcare in 2019 [109]. Early diagnosis of VHDs allows for timely therapeutic interventions such as regular screening checkups, lifestyle modifications, and medications to foster increased life expectancy among VHD patients [2]. According to a study conducted on 919 aortic stenosis patients during disease characterization and post-surgery observation periods, timely interventions were demonstrated to contribute to a decrease in the healthcare cost of VHD patients [3].

Initial evaluation of patients suspected with valve diseases is performed with the investigation of the patient's medical history along with the auscultation practice at the primary care physician's office. If suspicious heart murmurs are detected on the stethoscope, the patient is referred to a cardiologist for a transthoracic echocardiogram (TTE) test [2]. In the TTE test, a handheld ultrasound transducer is used to non-invasively create an image of the heart to establish valve morphology and cardiac chamber size, as

well as their impacts on the pulmonary system [4]. According to a study conducted on 213,174 VHD patients in 2019, it was demonstrated that a combination of a few TTE parameters such as aortic valve peak velocity, mean pressure gradient, and peak-velocity ratio could provide sensitivity and specificity values of 92% and 99% respectively for the diagnosis of aortic stenosis [5]. The achievement of such diagnostic power however, requires expertise in cardiac imaging and meticulous attention to the interpretation of hemodynamics [110]. Furthermore, lack of access to TTE hinders 52% of primary care physicians from confidently detecting VHDs through the use of stethoscopes [111]. This statistic is further confirmed by the low sensitivity of 44% reported for VHD detection based on the auscultatory examination of 251 patients presented in [6]. Hence, there is an imperative need for automatic, ubiquitous, and accurate detection of valvular heart diseases.

Following the recent advances in sensor technologies and artificial intelligence, the detection of valvular heart diseases through electrocardiogram (ECG), phonocardiogram (PCG), seismo-cardiogram (SCG), and gyro-cardiogram (GCG) has been investigated over the past decade. In [112], a DenseNet convolutional neural network (CNN) was tested on the 12-lead ECG data of 102,926 subjects and sensitivity, specificity, and area under the curve (AUC) of 78%, 74%, and 0.85, respectively, were reported for the identification of moderate to severe aortic stenosis (AS). In a similar study in [113], a combination of CNN and single-lead ECG data was demonstrated to predict AS with AUC, sensitivity, and specificity of 0.88, 0.90, and 0.71, respectively. The ValveNet deep neural network was recently introduced in [114] to classify patients into a wider range of valvular diseases

including aortic stenosis (AS), aortic regurgitation (AR), and mitral regurgitation (MR). As a result, AUC, sensitivity, and specificity of 0.88, 78%, and 73% were reported respectively. Despite the relatively acceptable performance, the use of electrodes and lead wires causes inconvenience for patients in clinical settings, thus limiting the use of ECG-based monitors as a ubiquitous technology.

In recent years, the recording of phonocardiogram (PCG), which signifies heart sounds, was proposed for the identification of valvular abnormalities such as aortic stenosis (AS), aortic regurgitation (AR), mitral stenosis (MS), mitral regurgitation (MR), pulmonary stenosis (PS), bicuspid aortic valve (BAV), and mitral valve prolapse (MVP). The authors in [115] proposed a time-growing neural network (TGNN) to characterize semilunar valve abnormalities using PCG recordings. They reported average accuracy, sensitivity, and specificity values of 84.2%, 82.8%, and 86.7%, respectively. In [116], [117], and [118], random forest (RF) classifiers were trained respectively on the features extracted from the time-frequency representation, instantaneous frequency, and Hilbert envelope of PCG recordings. Accuracies of 95.1%, 94.8%, and 98.8%, respectively, were achieved on the classification of valvular diseases. In [119] and [120], Mel-frequency cepstral coefficients (MFCCs) were extracted from PCG signals and trained on support vector machines (SVM), which resulted in accuracies of 92.6% and 97.9%, respectively, for the identification of valvular abnormalities. Recently, promising accuracies of 99.3% and 98.9% were reported respectively through the use of recurrent neural networks (RNNs) [121] and convolutional neural networks (CNNs) [122] for the detection of VHDs. It is to be noted that PCG is commonly recorded through a clinical stethoscope which is a less

complex device compared to ECG wires. However, PCG sensors are vulnerable to airborne noise leaking into the stethoscope chest piece and masking the heart sounds [123]. Furthermore, the sensitivity obtained from stethoscopes is as low as 60% for certain VHDs in patients with body mass index (BMI) values greater than 30 $kg/m^2$ [124]. This provokes the need for BMI-robust modalities.

Our research group has recently introduced SCG and GCG sensors to respectively investigate the linear and rotational chest vibrations associated with valvular heart diseases [71,125,126]. In [71], we demonstrated an accuracy of 96.4% for the classification of AS and the coexisting valvular diseases using wavelet features of 20 AS patients. Our further studies on 32 AS patients in [125] concluded that morphological characteristics of cardio-mechanical signals could characterize AS with high sensitivity, accuracy, and F1 score values of 97.3%, 96.4%, and 96.5% respectively. As confirmed in [71], SCG/GCG modalities could not perfectly characterize vibrational patterns of certain coexisting diseases, e.g. the sensitivity achieved for AS + TR was reported as 90%.

This research focuses on overcoming the limitations associated with PCG and SCG/GCG recordings for the detection of coexisting valvular heart diseases (VHD). The contributions of the work are summarized as follows:

- *To the best of our knowledge, this is the first study introducing contact microphones for the detection of coexisting valvular heart diseases.* The use of a sensitive contact microphone allows for the detection of minute heart-induced

vibrations on the chest wall, providing valuable information about the cardiac system.

- The concept of mapping from the 2-dimensional features of contact microphone recordings to a 5-digit binary sequence corresponding to 5 coexisting valvular heart diseases is introduced and formulated. Mel-frequency cepstral coefficients which imitate the human auditory system are proposed for the identification of coexisting VHDs.
- A novel image-to-sequence translation network based on the convolution-meets-transformer concept is proposed to solve the non-linear mapping from 2-dimensional representations to their corresponding 5-digit sequences. This architecture leverages convolutional and self-attention layers to characterize the local and global dependencies in ACM recordings, respectively.

### 2.2.1 Background on Heart Sounds and Abnormalities

The human cardiac cycle is signified by an audible rhythmic sound which is commonly referred to as the “lub-dub”. The “lub” component is the first heart sound ($S_1$) generated as a result of the closure of the mitral and tricuspid valves. The “dub” sound, which is the second heart sound ($S_2$), corresponds to the closure of the aortic and pulmonary valves. Hence, a cardiac cycle is denoted by the waveform between two successive $S_1$ sounds, with $S_1$-$S_2$ and $S_2$-$S_1$ marking systolic and diastolic periods respectively as illustrated in Fig. 2.11 (a).

In addition to $S_1$ and $S_2$, additional components may also exist in the heart sound signal of a valvular heart disease patient. These components are called murmurs, and may emerge

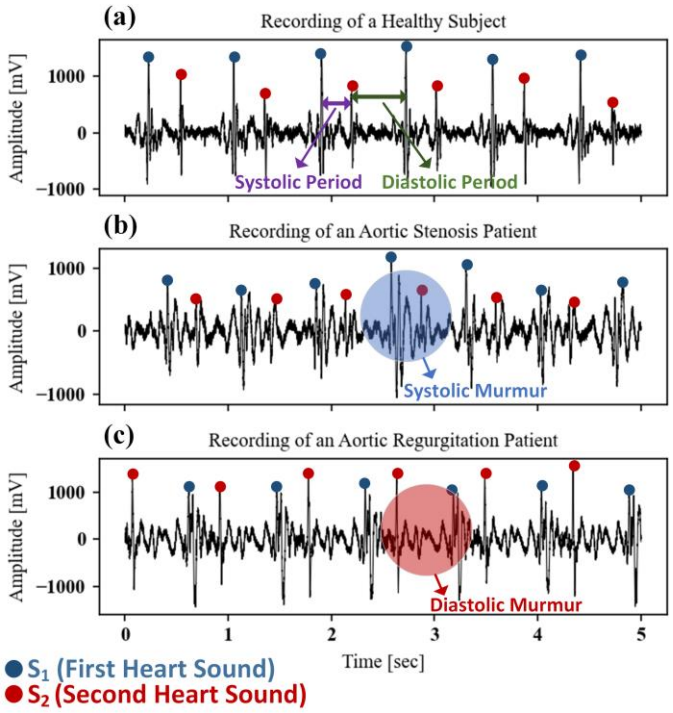


Fig. 2.11 Comparison of cardiac cycle morphologies in the ACM recordings of subjects with (a) a healthy cardiac system, (b) systolic murmurs, and (c) diastolic murmurs.

in different portions of the cardiac cycle depending on the type of valvular disease [127]. For instance, aortic stenosis (AS), mitral regurgitation (MR), and tricuspid regurgitation (TR) are associated with systolic murmurs, an example of which is shown in Fig. 2.11 (b), whereas other diseases such as aortic regurgitation (AR) and pulmonary regurgitation (PR) are associated with diastolic murmurs as shown in Fig. 2.11 (c). In addition to the occurrence interval of the murmurs, their shape should also be considered [118]. For instance, AS murmurs resemble diamonds in the systolic intervals of the cardiac cycle. The characterization of murmurs therefore, would allow for the identification of VHDs in a patient. This however is not feasible through auscultation as the coexistence of multiple diseases coupled with the low sensitivity of the human auditory system could inhibit physicians from accurately identifying VHDs [5]. In this work, we hypothesize that

automatic pattern recognition techniques are key to detecting coexisting VHDs in acoustic recordings of the cardiac system.

### 2.2.2 Experimental Setup and Methodology

#### 2.2.2.1 Data Collection and Study Protocol

As summarized in Table 2.8, this study was conducted on fifty-eight valvular heart disease (VHD) patients including 25 males and 33 females at the cardiac care unit of Sorin Medical P.C. in New York City. The VHD cohort is represented by average (± standard deviation) age, weight, and height values of 75.75 (± 12.51) years, 81.98 (± 19.05) kg, and 164.31 (± 8.98) cm, respectively. Two interventional cardiologists reviewed the patients' electrocardiogram (ECG) and echocardiogram (echo) records, and labeled them with one or a set of coexisting VHD diseases, following the American College of Cardiology/ American Heart Association (ACC/AHA) guidelines for the diagnosis and management of

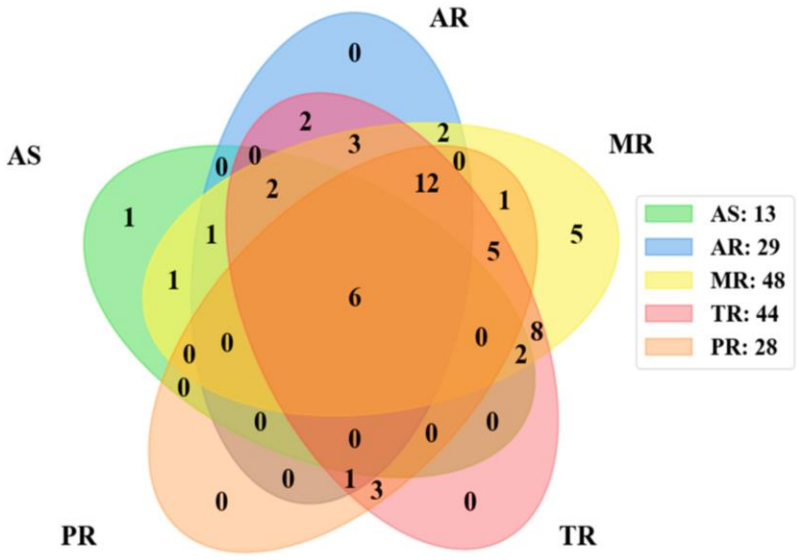


Fig. 2.12 The distribution of the VHD patients in terms of their coexisting valvular diseases.

VHD patients. Fig. 2.12 depicts a Venn diagram corresponding to the distribution of the diseases among the patient cohort. As can be observed, the VHDs studied in this study include aortic stenosis (AS), aortic regurgitation (AR), mitral regurgitation (MR), pulmonary regurgitation (PR), and tricuspid regurgitation (TR). The intersections among the sets in Fig. 2.12 imply the presence of multiple VHDs in certain patients. *In this study, each patient is assigned a 5-digit binary sequence representing their coexisting VHDs*, for instance "10100" for a patient with AS and MR and no other VHDs. Additionally, a control group consisting of 21 healthy males and 31 healthy females with average (± standard deviation) age, weight, and height of 66.59 (± 12.55) years, 76.80 (± 20.46) kg, and 164.35 (± 10.73) cm were measured for comparison with the VHD cohort. It is to be mentioned that sequences associated with the healthy subjects are denoted as "00000", conveying the absence of all types of VHDs.

Fig. 2.13 illustrates the experimental setup for the detection of coexisting valvular heart diseases. As shown in Fig. 2.13 (a), a $\pm 4$ g sensitive accelerometer contact microphone (ACM) with micro-g resolution [128,129] (obtained from StethX Microsystems Inc., Atlanta, USA) is attached to the chest wall along the third rib using medical-grade adhesive tape. Additionally, the experimental setup incorporates a photoplethysmography (PPG) sensor (Shimmer3 GSR; Shimmer Sensing, Dublin, Ireland) clipped to the ear-lobe, as well as a three-lead ECG sensor node (ECG Development Kit; Shimmer Sensing, Dublin, Ireland) with the electrodes attached to the right arm (RA), left arm (LA), right leg (RL), and left leg (LL). The ACM sensor head has a small form factor of 27 mm × 15 mm × 2.5 mm which is comparable to the size of a dime as depicted in Fig. 2.13 (b). This device is a

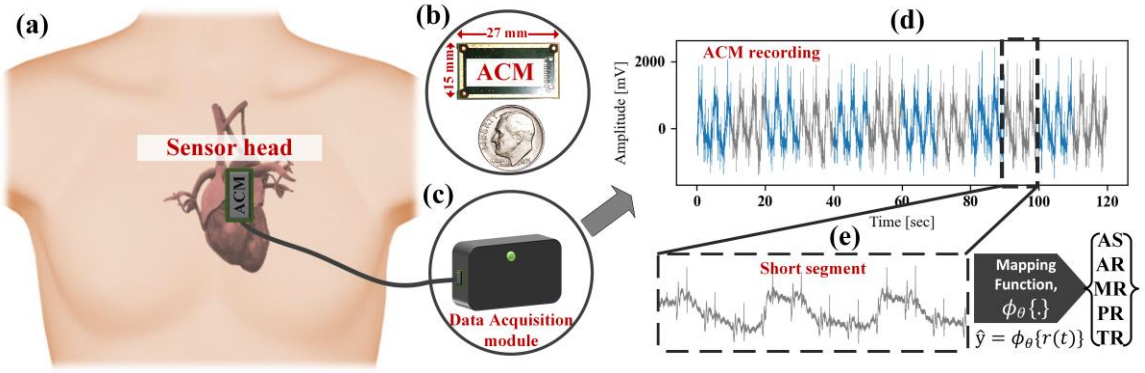


Fig. 2.13 The experimental setup for the detection of coexisting valvular heart diseases. (a) The sensor head non-invasively attached to the chest wall using adhesive tape, (b) the size of the ACM sensor head compared to a dime, (c) the data acquisition module used for data transmission, (d) a two-minute ACM recording, and (e) a 10-second signal segment used to predict the sequence associated with coexisting VHDs.

low-noise accelerometer with a wide operational bandwidth of 0-10 kHz, allowing for recording heartbeat-induced sounds and vibrations on the chest wall. The sensor is not sensitive to airborne emission sounds, making it a robust phonocardiogram sensor against environmental acoustic noise. The recorded data is digitized using a 24-bit analog-to-digital converter integrated into the sensor head, and transferred to a computer in real-time through a data transfer cable connected to the data acquisition (DAQ) module, as shown in Fig. 2.13 (c). The subjects were seated at rest on a chair for a period of five minutes, followed by five minutes of measurement with ACM at a sampling rate of 22.33 kHz, still at a seated position. The patient experimental protocol was approved by the Institutional Review Board of Stevens Institute of Technology under protocol number 2022-044 (N). It is to be noted that ECG and PPG data are not used in this specific study. In the proposed framework, the objective is to predict sequences associated with coexisting VHDs using a short segment of ACM recordings as shown in Figs 2.13 (d) and (e).

#### 2.2.2.2 VHD Detection Problem Statement

As earlier, valvular abnormalities appear as irregular shapes in ACM recordings. As opposed to previous studies which primarily aim to relate each patient's recording to only a specific type of valve disease, we formulate the problem for the detection of coexisting VHDs. For this purpose, we model the heart sound of a VHD patient as follows:

$$r(t) = \psi(s(t), m(t)), \qquad (7)$$

where $r(t)$, $s(t)$, and $m(t)$ denote the chest recording, heart sounds, and murmurs, respectively, and $\psi(.)$ represents a function comprised of mixed heart sound and murmur components. To determine the coexisting disease type, we need to determine murmur components. To this end, we define a transformation, i.e., $\phi_\theta\{.\}$ with $\theta$ denoting the transformation parameters, to map $r(t)$ to a 5-digit binary sequence $\mathrm{y} = (\mathrm{AS}, \mathrm{AR}, \mathrm{MR}, \mathrm{PR}, \mathrm{TR})$. Hence:

$$\hat{\mathrm{y}} = \phi_\theta\{r(t)\}, \qquad (8)$$

with $\hat{\mathrm{y}}$ representing the estimated binary sequence corresponding to $r(t)$. An optimization problem is defined as follows to determine the optimum parameters, i.e., $\theta^*$, minimizing the distance between y and $\hat{\mathrm{y}}$:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} |\mathrm{y} - \phi_\theta\{r(t)\}|. \qquad (9)$$

This optimization signifies the objective of this study, as described in the following sections.

#### 2.2.2.3 Data Preparation

In order to accelerate the processing chain, ACM recordings were initially downsampled to 4 kHz prior to further analysis. This practice does not cause information loss since heart-induced components fall below the frequency of 1 kHz [130]. The ACM signals were then segmented into 10-second overlapping windows with 1-second sliding steps. The choice of the window length follows the superior results obtained for the 10-second duration reported in [125]. The preprocessing step begins with zero-padding the beginning and end of each segment by 256 points (62.5 ms) to prevent the transient effect of filtering performed in the following steps. As illustrated in Fig. 2.13 (e), heart sound components are represented by small peaks modulated on the low-frequency respiration waveform. A differentiation filter was applied to the signal segment to amplify high-frequency components associated with heart sounds which were then high-pass filtered using a 3$^{rd}$-order Butterworth filter with a cut-off frequency of 5 Hz. Subsequently, a cumulative filter was used to reconstruct the heart sound morphology, and the padded points at the beginning and end of each segment were discarded to restore a clean signal with the same length as the original signal. To further enhance the signal-to-noise ratio (SNR) of the ACM recordings, a denoising algorithm was employed to alleviate the impact of noise components. Several studies have demonstrated the potential of empirical mode decomposition (EMD) for phonocardiogram denoising in the frequency domain [44] or time domain [131,132]. For this purpose, we adopt the denoising algorithm proposed in [132]. This method decomposes each signal segment into intrinsic mode functions (IMFs) as follows:

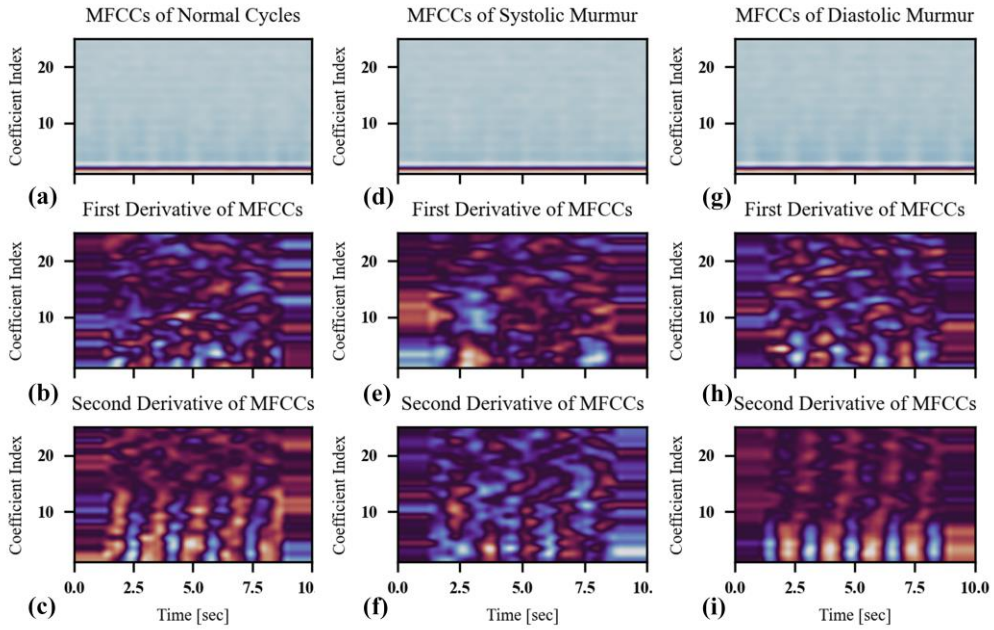


Fig. 2.14 Mel-frequency cepstral coefficients (MFCCs) and their first and second derivatives for normal cycles (a-c), systolic murmur-contaminated cycles (d-f), and diastolic murmur-contaminated cycles (g-i), respectively.

$$r(n) = \sum_{i=1}^{L-1} \mathrm{IMF_i}(n) + \mathrm{R}(n), \qquad (10)$$

where $\mathrm{IMF_i}(n)$, $L-1$, and $\mathrm{R}(n)$ denote the $i^{th}$ IMF, number of IMFs, and the final residue signal respectively. Then according to [132], the actual energy of each IMF is compared to its estimated energy, and if the actual energy exceeds the estimated one, the IMF is included for signal reconstruction. Hence, the denoised signal, i.e., $r'(n)$ is reconstructed by the summation of all eligible IMFs, i.e., $\widehat{\mathrm{IMF}(n)}$, as follows:

$$r'(n) = \sum_{i=1}^{M} \widehat{\mathrm{IMF_i}}(n), \qquad (11)$$

with $M$ denoting the number of relevant IMFs. Finally, the signal is passed through a 3rd-order Savitzky-Golay smoothing filter of size 50 ms to produce smooth outputs while

maintaining signal patterns [133]. Each 10-second denoised signal is then used for feature extraction, pattern recognition, and decision-making as described in the following sections.

#### 2.2.2.4 Feature Extraction

As described earlier, valvular heart diseases are represented by murmurs in the heart sounds. In cardiac auscultation, a cardiologist uses a stethoscope to listen to heart sounds and distinguish abnormal auditory patterns. Inspired by auscultation, we design our feature extraction framework to imitate the human auditory system.

Mel-frequency cepstral coefficients are a set of features that distribute the short-term power spectrum of a signal in accordance with the human auditory system perception in a scale called Mel [134]. This scale employs a filter spaced linearly at frequencies below 1 kHz and logarithmically above 1 kHz. MFCCs computation begins with framing a signal segment into shorter frames of 0.5 seconds (2,048 samples) with 25% (512 samples) overlap between consecutive frames. We employed a Mel filter bank with 25 triangular filters to extract 25 coefficients associated with the periodogram of each frame. Finally, the discrete cosine transform is applied to the coefficients to obtain MFCCs. Hence, each 10-second signal segment is represented by a 2-D representation of size 25 × 81, where 81 represents the number of frames and 25 denotes the number of coefficients per frame. Figs 2.14 (a), (d), and (g) depict MFCCs corresponding to heart rhythms with a normal cycle, systolic murmur, and diastolic murmur respectively. In order to further model the dynamics of the heart sounds, the first and second derivatives of MFCCs are calculated by subtracting the coefficients of successive time frames, once for the first derivative and twice for the second derivative as shown in Figs 2.14 (b-c), (e-f), and (h-i) for heart rhythms with

normal, systolic murmur, and diastolic murmur components respectively. Although visual inspection suggests distinctive patterns for each condition, the detection of coexisting patterns would be challenging unless an automatic pattern recognition model is employed.

#### 2.2.2.5 Image-to-Sequence Translation

As shown in Fig. 2.14, abnormal signal patterns are visible in MFCCs and their derivative features. In this section, we design a neural network to solve the optimization problem described in (9). Accordingly, the desire is to map the extracted features for each 10-second segment to its corresponding sequence, e.g., “10100”, which implies an image-to-sequence translation task. In recent years, convolutional neural networks (CNNs) have made significant contributions to computer vision tasks, such as image classification [135,136], denoising [137], lung sound auscultation [138], and segmentation [139–141] due to their ability to extract discriminative features. However, CNNs suggest limited performance for modeling long-range dependencies due to the small receptive fields of convolutional kernels [142]. To address this limitation, transformers which were originally proposed for sequence modeling were adapted to computer vision tasks to find the relevance among the patches constituting an image [142]. Despite the excellent performance at sequence modeling [143], transformers show inferior performance in computer vision tasks compared to similar-sized CNNs since patch-based modeling withholds transformers from encoding local information. In [132], we demonstrated that increasing the number of patches through transfer learning could result in superior performance compared to transformer- or CNN-alone architectures.

In this study, we adopt the "convolution-meets-transformer (CMT)" concept proposed by [144], and develop an architecture that leverages the advantages of both CNNs and transformers for extracting local and global information in an image respectively, as depicted in Fig. 2.15. This combination allows us to relate maps of MFCCs to their corresponding sequences, i.e., image-to-sequence translation. As mentioned earlier, the

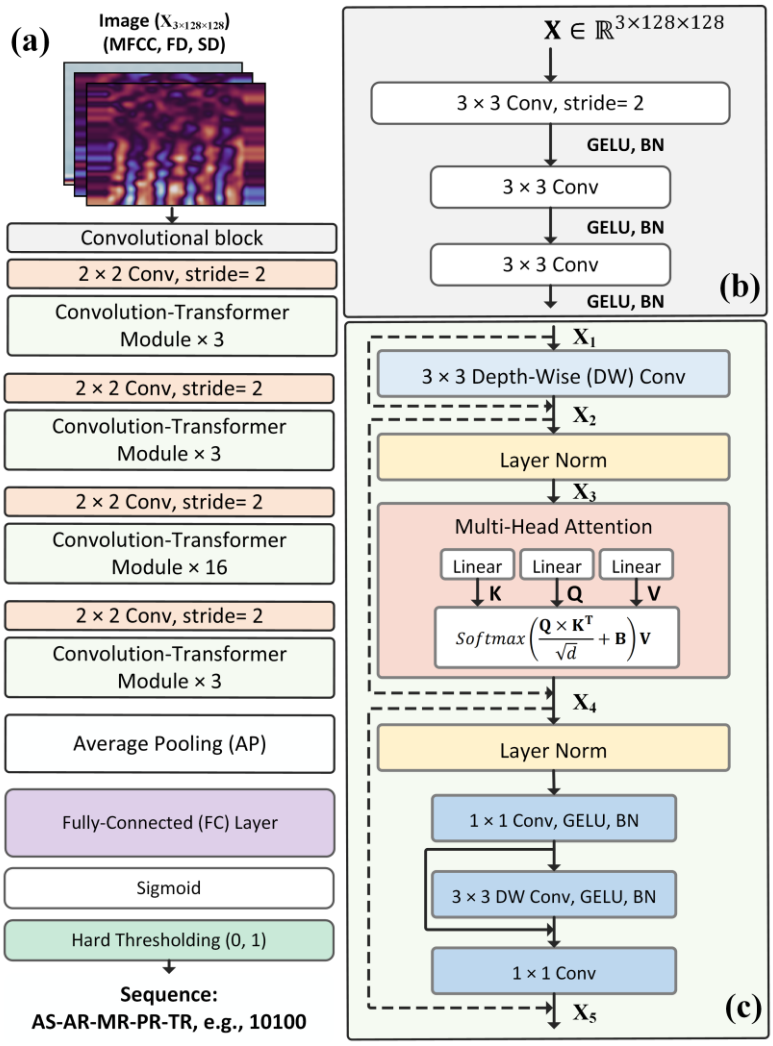


Fig. 2.15 (a) the image-to-sequence translation network developed in this work for the detection of coexisting valvular heart diseases, (b) the convolutional block consisting of three convolutional layers, and (c) the convolution-transformer module (CTM).

dimensions of MFCCs, the first derivative (FD), and the second derivative (SD) are 25 × 81. We initially up-sample MFCCs and the corresponding first and second derivatives through a bilinear interpolation to 128 × 128 to have standard-size inputs ($\mathbf{X} \in \mathbb{R}^{3\times128\times128}$). As illustrated in Fig. 2.15 (a), **X** undergoes a convolutional block which is detailed in Fig. 2.15 (b). This layer is composed of a 3×3 convolutional layer with 8 kernels and a stride of 2, followed by Gaussian error linear unit (GELU) [145] and batch normalization (BN) [146] for non-linear modeling and training stabilization respectively, to reduce the size of the input image while extracting 8 feature maps. Two 3×3 convolutional layers with 32 channels and strides of 1 are then applied to further extract fine-grained patterns observed in contiguous bins in the MFCC-time plane.

As shown in Fig. 2.15 (a), four similar stages, each including a 2×2 convolutional layer with a stride of 2 followed by a convolution-transformer module (CTM), are considered to extract local and global features. 2×2 convolutional layers are employed to increase the feature dimension and reduce the input resolution both by a factor of 2, resulting in a 16×32×32 feature set prior to the first CTM. The architecture of CTM is depicted in Fig. 2.15 (c), where a 3×3 depth-wise (DW) convolutional layer is applied to the input feature set $X_1$, and generates the output as follows:

$$\mathbf{X_2} = \mathbf{X_1} + \mathrm{DW}(\mathbf{X_1}). \qquad (12)$$

In the next step, a layer normalization (Layer Norm) is utilized to ensure a fast convergence, as follows:

$$\mathbf{X_3} = \frac{\mathbf{X_2} - \mathrm{E}[\mathbf{X_2}]}{\sqrt{\mathrm{Var}[\mathbf{X_2}] + \epsilon}} \times \gamma + \beta, \qquad (13)$$

with E[.], Var[.] signifying mean and variance operators respectively, and $\epsilon$, $\gamma$, and $\beta$ denoting normalization learnable parameters. Considering $\mathbf{X_3} \in \mathbb{R}^{C\times H\times W}$ as $n = H \times W$ image patches, i.e., elements, with length $C$, the attention mechanism multiplies each element with three learnable matrices $W^Q \in \mathbb{R}^{d\times C}$, $W^K \in \mathbb{R}^{d\times C}$, and $W^V \in \mathbb{R}^{d\times C}$. This produces three vectors called query ($q_i \in \mathbb{R}^{1\times d}$), key ($k_i \in \mathbb{R}^{1\times d}$), and value ($v_i \in \mathbb{R}^{1\times d}$) for each element. The query of each element is multiplied by all the keys in the sequence to find the relation between the element under test and other elements in the sequence. Additionally, we add a learnable matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ to the attention module to encode the relative positions of elements. The score vector which is of size $n$ subsequently undergoes a Softmax function to rescale the scores into the range of [0, 1], and make them sum up to 1. These scores are multiplied by their respective value vectors to build the attention module output for the element under test. Hence, the attention module can be formulated as:

$$\mathbf{U} = Softmax\left(\frac{\mathbf{Q}\times\mathbf{K}^{\mathrm{T}}}{\sqrt{d}}\right)\mathbf{V}, \qquad (14)$$

where $Q \in \mathbb{R}^{n\times d}$, $K \in \mathbb{R}^{n\times d}$, and $V \in \mathbb{R}^{n\times d}$ denote the matrices built by stacking query vectors, key vectors, and value vectors respectively.

Matrix U $\in \mathbb{R}^{n\times d}$ shows the output matrix with $n$ rows corresponding to patches each with depth $d$. In order to have a robust attention mechanism and thus better performance, the use of multiple heads is suggested by [143], where the output of all attention heads is averaged. In this work, the number of attention heads is 1, 2, 4, and 4 for CMT 1, 2, 3, and 4 respectively. Therefore, the new feature set is created as:

$$\mathbf{X_4} = \mathbf{U} + \mathbf{X_2}, \qquad (15)$$

where $X_4 \in \mathbb{R}^{d \times H \times W}$, and $d$ denotes the embedding size, implying an increase in the feature channels by a factor of $d/C$. It is to be noted that $d$ is set to 16, 32, 32, and 64 for CTM 1, 2, 3, and 4 respectively. Subsequent to Layer Norm, a point-wise convolutional layer (i.e. 1×1) is applied to $X_4$ to expand the number of channels by a factor of 4. This is then followed by depth-wise and point-wise convolutional layers to extract local features and shrink the number of channels to $C$. This creates the output of CTM which is represented by $X_5 \in \mathbb{R}^{128 \times 4 \times 4}$. As suggested by [144], we used 3 CTMs in stages 1, 2, and 4, and 16 CTMs in stage 3 to extract features from MFCCs features, as shown in Fig. 2.15 (a). The logic behind such a hierarchical architecture is to generate feature maps of different scales and levels of abstraction.

In order to estimate the VHD sequence, the information of all 128 channels is globally averaged within the feature map, making each image patch contribute to the estimation of the VHD sequence. A fully-connected (FC) layer with 5 neurons followed by a sigmoid function is then applied to the averaged values to produce values within the range of [0,1] for each VHD disease type. Finally, a hard thresholding function is applied to create a binary sequence from the sigmoid output. This function replaces neurons less than 0.5 with 0, and values greater than 0.5 with 1. As such, an input image consisting of MFCCs, FD, and SD of an ACM recording is translated to the corresponding binary sequence of coexisting VHDs. The objective of this network is to minimize the mean square error (MSE) between the predicted VHD sequence, i.e., $y_i, i \in 1, 2, 3, \ldots, N$, and the

corresponding ground-truth sequences, i.e., $\hat{y}_i, i \in 1, 2, 3, \dots, N$ with $N$ denoting the total number of test samples, as defined below:

$$\ell(y, \hat{y}) = \frac{1}{N}\sum_{i=0}^{N} |y_i - \hat{y}_i| \,, \qquad (16)$$

where $\ell(y, \hat{y})$ denotes the objective function.

### 2.2.3 Experimental Results

In this section, an overview of the training procedure, the evaluation metrics, and the experimental results are described.

#### 2.2.3.1 Training Procedure

As described earlier, each signal segment is converted to a three-channel image consisting of MFCCs, FD, and SD, which are used as the input to the CMT network shown in Fig. 2.15 (a). This network is trained in a supervised manner to translate the input images to their corresponding VHD sequences. The training procedure was conducted using the data of 110 subjects amounting to 36,939 signal segments. In this work, the generalizability of CMT to unseen data was assessed based on a 10-fold leave-subject-out cross-validation (10-LSOCV) approach. For this purpose, the dataset was split into the train (77 subjects; 70%), validation (22 subjects; 20%), and test (11 subjects; 10%) sets. The image-to-sequence translation network was implemented on a 24-GB NVIDIA GeForce RTX 3090 Ti FTW3, and trained on the train set with a batch size of 64 for 100 epochs. The Adam optimization algorithm [147] with an initial learning rate of 0.005 was employed to minimize the loss function, defined in (10), with respect to the network parameters. At each

Table 2.9 10-LSOCV Average (± Standard Deviation) Performance on VHD Detection for the Validation Dataset

| Disease | Sensitivity (%) | Specificity (%) | Accuracy (%) | PPV (%) | F1 score (%) |
|---|---|---|---|---|---|
| **Aortic Stenosis** | 99.05 (± 1.25) | 99.64 (± 0.26) | 99.60 (± 0.29) | 95.42 (± 3.33) | 97.18 (± 2.06) |
| **Aortic Regurgitation** | 95.50 (± 4.48) | 97.17 (± 1.40) | 96.96 (± 1.07) | 84.22 (± 6.52) | 89.28 (± 3.40) |
| **Mitral Regurgitation** | 96.81 (± 1.65) | 98.78 (± 1.49) | 97.84 (± 1.40) | 98.63 (± 1.67) | 97.70 (± 1.49) |
| **Tricuspid Regurgitation** | 97.13 (± 2.09) | 98.65 (± 1.51) | 98.08 (± 1.46) | 97.73 (± 2.50) | 97.41 (± 1.96) |
| **Pulmonary Regurgitation** | 94.31 (± 3.95) | 99.45 (± 0.54) | 98.45 (± 0.97) | 97.67 (± 2.31) | 95.92 (± 2.60) |
| **Aggregate (± Std.)** | **96.56 (± 3.28)** | **98.74 (± 1.42)** | **98.19 (± 1.38)** | **94.73 (± 6.46)** | **95.50 (± 3.93)** |

fold, the model was validated on the validation set and the performance was reported using the test set when the validation loss was minimized.

### 2.2.3.2 Evaluation Metrics

Statistical analyses were performed to evaluate the performance of the CMT network and compare it with state-of-the-art VHD detection methods. In order to compare the ground-truth sequences with those translated by CMT, sensitivity (Sen), specificity (Spec), accuracy (Acc), positive predictive value (PPV), and F1 score (F1) values were calculated as follows:

$$\text{Sen} = 100 \times \frac{TP}{TP+FN}, \qquad (17)$$

$$\text{Spec} = 100 \times \frac{TN}{TN+FP}, \qquad (18)$$

$$\text{Acc} = 100 \times \frac{TP+TN}{TP+TN+FP+FN}, \qquad (19)$$

$$\text{PPV} = 100 \times \frac{TP}{TP+FP}, \qquad (20)$$

$$\mathrm{F1} = \frac{2\times \mathrm{PPV}\times \mathrm{Sen}}{\mathrm{PPV}+\mathrm{Sen}}, \qquad (21)$$

where $TP$, $TN$, $FP$, and $FN$ signify true positive (correctly identified as VHD), true negative (correctly identified as healthy), false positive (falsely identified as VHD), and false negative (falsely identified as healthy) cases, respectively. Sensitivity implies the rate of falsely-missed VHD cases with respect to the correctly-detected VHD cases, whereas specificity indicates falsely-detected VHD cases with respect to truly-identified healthy subjects. Positive predictive value signifies the rate of truly-detected VHD cases with respect to the rate of missed VHD cases, implying the likelihood that CMT could detect VHD in a patient. The F1 score was adopted to present a harmonic mean of positive predictive value and sensitivity of CMT, to quantify the balance between the rate of true detection of VHDs compared to false decisions, i.e., $FP$ and $FN$.

#### 2.2.3.3 VHD Detection Performance

Most of the prior arts on VHD detection have reported their predictive performance only when the validation loss is minimized at each fold. This however, does not imply a fully

Table 2.10 10-LSOCV Average (± Standard Deviation) Performance on VHD Detection for the Test Dataset

| Disease | Sensitivity (%) | Specificity (%) | Accuracy (%) | PPV (%) | F1 score (%) |
|---|---|---|---|---|---|
| **Aortic Stenosis** | 94.93 (± 5.12) | 99.33 (± 0.49) | 98.97 (± 0.64) | 93.52 (± 4.23) | 94.14 (± 3.75) |
| **Aortic Regurgitation** | 93.11 (± 5.97) | 96.08 (± 1.66) | 95.63 (± 1.17) | 82.07 (± 6.14) | 86.96 (± 3.37) |
| **Mitral Regurgitation** | 94.41 (± 2.88) | 97.81 (± 2.55) | 96.13 (± 2.53) | 97.36 (± 3.44) | 95.85 (± 2.90) |
| **Tricuspid Regurgitation** | 94.26 (± 3.59) | 98.3 (± 2.00) | 96.63 (± 2.51) | 97.55 (± 2.96) | 95.86 (± 3.09) |
| **Pulmonary Regurgitation** | 89.70 (± 5.19) | 98.78 (± 0.89) | 97.01 (± 1.38) | 94.34 (± 4.66) | 91.86 (± 3.87) |
| **Aggregate (± Std.)** | **93.28 (± 4.88)** | **98.07 (± 1.97)** | **96.87 (± 2.08)** | **92.97 (± 7.13)** | **92.94 (± 4.69)** |

unbiased evaluation of the predictive model since the lowest loss is always considered in this approach. On the other hand, a test set, on which the loss is not necessarily minimized, could quantify the generalizability of the trained model to unseen data. Hence, we evaluate the performance using both validation and test sets to follow the conventions and provide an unbiased assessment of VHD detection. Statistical analyses on the validation and test sets are presented in Tables 2.9 and 2.10, respectively, for 10-LSOCV. As mentioned in Table 2.9, the highest average (± standard deviation) sensitivity of 99.05% (± 1.25) was achieved for the detection of AS, implying a more evident presentation in ACM recordings compared to AR, MR, TR, and PR with average (± standard deviation) sensitivities of 95.50% (± 4.48), 96.81% (± 1.65), 97.13% (± 2.09), and 94.31% (± 3.95) respectively, whereas the lowest sensitivity belongs to PR. These results are consistent with the sensitivities of 94.93% (± 5.12) and 89.70% (± 5.19) reported for AS and PR in Table 2.10 respectively, as the highest and lowest performances. As suggested by Table 2.9, the highest and lowest specificities were achieved for AS with 99.64% (± 0.26) and AR with 97.17% (± 1.40) respectively. This implies higher false positive rates in AR compared to other VHDs, thus leading to more falsely-detected AR cases. Comparing the specificities in Table 2.10 with Table 2.9, the same trend is observed among VHDs when moving downwards from AS to PR, where the highest and lowest specificities of 99.33% (± 0.49) and 96.08% (± 1.66) are reported for AS and AR, respectively, which agrees with the findings from the validation set.

As mentioned in Tables 2.9 and 2.10, the average accuracies fall within the ranges of [96.96%, 99.60%] and [95.63%, 98.97%] for the validation and test sets respectively.

According to the aggregate accuracy scores, the prediction errors are estimated to be as low as 1.89% and 3.13% for the validation and test sets, respectively, suggesting high consistencies between the ground-truth and predicted sequences. Relatively low positive predictive values of 84.22% (± 6.52) and 82.07% (± 6.14) reported for AR respectively in Tables 2.9 and 2.10 indicate higher falsely-detected AR cases compared to correctly-identified cases, which directs healthy cases for further diagnosis process. The highest and the second-highest positive predictive values of 98.63% (± 1.67) and 97.36% (± 3.44) belong to MR, as mentioned in Tables 2.9 and 2.10 respectively. Tables 2.9 and 2.10 report the highest F1 scores of 97.70% (± 1.49) and 95.86% (± 3.09) for MR in the validation set and TR in the test set respectively, whereas the lowest F1 scores belong to MR in both validation and test sets with average scores of 89.28% (± 3.40) and 86.96% (± 3.37)

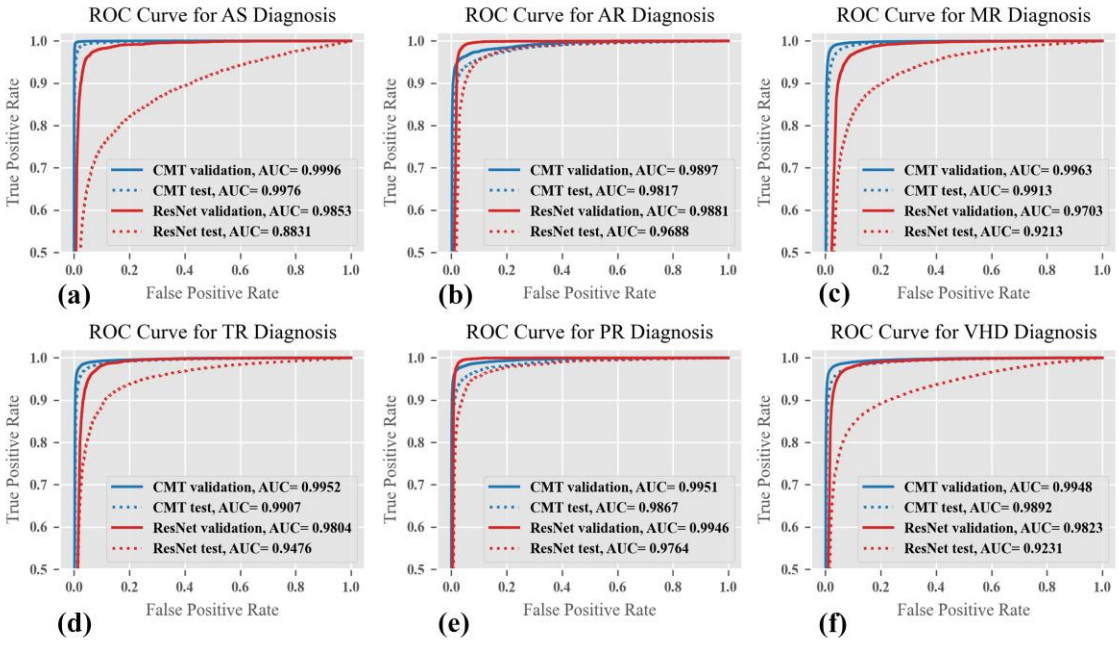


Fig. 2.16 Receiver operating characteristic curves for the diagnosis of (a) aortic stenosis, (b) aortic regurgitation, (c) mitral regurgitation, (d) tricuspid regurgitation, (e) pulmonary regurgitation, and (f) aggregate valvular abnormalities.

respectively. This finding confirms the low positive predictive values reported for MR (i.e. 84.22% and 86.96% for the validation and test sets respectively).

Additionally, we have reported the aggregate statistical scores in the last rows of Tables 2.9 and 2.10. As can be observed, aggregate sensitivity, specificity, accuracy, positive predictive value, and F1 score of 96.56% (± 3.28), 98.74% (± 1.42), 98.19% (± 1.38), 94.73% (± 6.46), and 95.50% (± 3.93) respectively were achieved for the validation set in Table 2.9, implying an excellent predictive power for the detection of coexisting VHDs based on the recordings acquired from the ACM. As for the test set, slightly lower aggregate sensitivity, specificity, accuracy, positive predictive value, and F1 score of 93.28% (± 4.88), 98.07% (± 1.97), 96.87% (± 2.08), 92.97% (± 7.13), and 92.94% (± 4.69) were obtained respectively, as suggested in Table 2.10. The superior performance of the validation set to that of the test set is expected since the model with the lowest validation loss does not necessarily have the lowest test loss. Nevertheless, we will demonstrate that the performance of CMT on the test dataset is as excellent as the conventional methods in the following sections.

#### 2.2.3.4 Receiver Operating Characteristic (ROC)

The diagnostic ability of the proposed framework is further evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) for each disease. This characteristic presents the true positive rate in terms of false positive rates in a binary manner as illustrated in Fig. 2.16. For each class, we compared the corresponding element in the ground-truth sequence with that in the predicted sequence, and determined the ROC curve corresponding to AS, AR, MR, TR, and PR as shown in Fig. 2.16 (a), (b), (c), (d),

and (e) respectively. Additionally, an aggregate ROC curve consisting of all the elements in the VHD sequence is shown in Fig. 2.16 (f). It is to be noted that ROC curves are included for both validation and test sets separately to be able to compare the proposed method with state-of-the-art methods. For the sake of comparison with off-the-shelf convolutional neural networks, the prediction results from a ResNet50 trained on the features of ACM recording are included in each ROC curve in Fig. 2.16. Also to be noted is that all ROCs are plotted for true positive rate greater than 0.5 to be able to show relative trends on the top-left corner of the curves.

According to Fig. 2.16, the AUCs of 0.9996, 0.9897, 0.9963, 0.9952, and 0.9951 were achieved by CMT on the validation sets for the detection of AS, AR, MR, PR, and TR respectively, with AS and AR holding the highest and lowest AUCs. The ROCs corresponding to the test sets evaluated on CMT suggest slightly smaller AUCs of 0.9976, 0.9817, 0.9913, 0.9907, and 0.9867 for AS, AR, MR, PR, and TR, respectively, as compared with the validation sets. As demonstrated in Fig. 2.16, the predictive robustness of ResNet50 is inferior to CMT for AS (0.8831 vs. 0.9976), AR (0.9688 vs. 0.9817), MR (0.9213 vs. 0.9913), TR (0.9476 vs. 0.9907), and PR (0.9764 vs. 0.9867) on the test sets. As depicted in Fig. 2.16 (f), the AUC of CMT on the aggregate of all VHD signal segments was reported as 0.9948, 0.9892, 0.9823, and 0.9231 for CMT + validation set, CMT + test set, ResNet + validation set, and ResNet + test set respectively, which implies an overall superior performance for CMT compared to ResNet. Such excellent AUCs coupled with the aggregate F1 score of 92.94% suggest the reliability of the proposed framework for the clinical detection of VHDs.

### 2.2.4 Discussion

This section discusses the robustness of MFCC + CMT framework against body mass index (BMI) variations, and presents a comprehensive comparison with state-of-the-art methods of VHD detection.

#### 2.2.4.1 Robustness Against BMI Variation

The body mass index (BMI), defined as the weight divided by the square of the height, describes the relative mass and height of an individual. People are classified based on their BMI into underweight ($\text{BMI} < 18.5\ kg/m^2$), normal ($18.5 < \text{BMI} < 24.9\ kg/m^2$), overweight ($18.5 < \text{BMI} < 24.9\ kg/m^2$), obese ($25 < \text{BMI} < 29.9\ kg/m^2$), and extra-obese ($\text{BMI} > 30\ kg/m^2$). As reported in [148], obesity contributes directly to the incidence of cardiovascular risk factors, including hypertension, diabetes, and sleep disorders. In a recent study on 108,211 individuals, it was demonstrated that BMI is highly correlated with the occurrence of valvular diseases such as aortic stenosis [149]. On the other hand, the low sensitivity of 60% obtained from stethoscopes for monitoring high-BMI VHD patients cause the disease to remain undetected as mentioned earlier. As such, the reliability of the proposed framework should be evaluated against BMI variations. Among the 110 recruited subjects in this study, 30 subjects had BMIs in the normal range, whereas 33, 22, and 16 subjects were identified as overweight, obese, and extra-obese respectively. For each BMI group, sensitivity, specificity, and F1 score of the classification results were determined as illustrated in Fig. 2.17. As can be seen, sensitivity scores of 96.22%, 95.98%, 93.29%, and 91.40% were reported for normal, overweight, obese, and extra-obese groups, respectively. In parallel, the specificity (98.59%, 97.79%, 97.60%, and

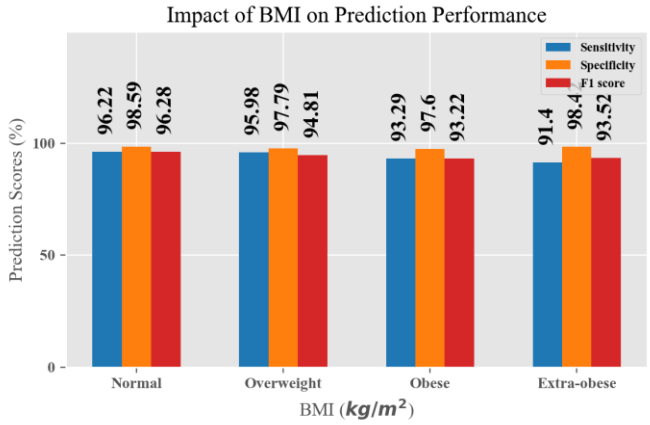


Fig. 2.17 The impact of body mass index (BMI) on the prediction scores for the detection of VHDs.

98.42% respectively) and F1 scores (96.28%, 94.81%, 93.22%, and 93.52% respectively) follow the same decreasing trend as the sensitivity score, implying slightly weaker predictive power for higher BMIs. The slight decrease in prediction scores from normal to extra-obese BMI can be explained by the fat tissues in high-BMI individuals which inhibit the heart-induced sounds and vibrations to reach the chest wall. Nonetheless, the small difference between the average F1 scores of normal and extra-obese groups (96.28% vs. 93.52%) demonstrates the robustness of the proposed VHD detection framework against BMI variations.

### 2.2.4.2 Comparative Analysis

The use of a sensitive accelerometer contact microphone in conjunction with a convolution-transformer network has allowed the detection of coexisting valvular diseases, a capability lacking from prior arts in the literature. Nonetheless, a comparative analysis of the prediction power between the proposed framework and state-of-the-art methods was conducted and presented in Table 2.11. It is worth mentioning that two types of validation

approaches including cross-validation (CV) and the use of test set were adopted by the literature. The former method splits the data into two parts, i.e., the train and validation sets, and reports the maximum performance on the validation set, whereas the latter splits the data into train, validation, and test sets, and the performance is reported for the test set using the model parameters leading to the lowest validation loss. Hence, the test set method provides a more meaningful metric of the generalizability of a predictive model in real-world scenarios. The third column in Table 2.11 describes the VHD(s) based on which the predictive scores are reported. All methods listed in the table, except for [71], are limited to the detection of either one type of VHD (binary classification), or only one out of many diseases (multi-class classification, but not coexisting). The random forest (RF) model driven by SCG/GCG in [71] simplified a multi-label classification task into a single-label multiclass classification by combining AS with other coexisting diseases, thus generating paired labels (AS-MR, AS-MS, and AS-TR). This inhibits the RF classifier from detecting more than two coexisting VHDs. Moreover, the identification of MR-only, MS-only, and TS-only patients remains unsolved. The fourth column in the table signifies the heartbeat segmentation strategies employed by the VHD detection methods. These strategies are based on either the cardiac cycle or X-second with X denoting the count of seconds. In the cardiac cycle segmentation method, the start and end of each heartbeat are annotated, and thus a single or a few heart cycles are used for further analysis and training of the predictive model. Cycle-based segmentation builds morphologically consistent training samples which enable the predictive model to better learn the distribution of the data. The annotation of PCG and SCG/GCG signals however, is generally not a straightforward task

Table 2.11 Comparative Analysis of VHD Detection Based on Accelerometer Contact Microphone Data Trained on the CMT Architecture Versus State-of-the-Art

| Methodology (modality) | Evaluation scheme | Disease(s) | Segmentation | Performance | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Sen (%) | Spec (%) | Acc (%) | F1 score (%) | AUC |
| DenseNet (ECG) [112] | Test set | AS | 10-second | 78.1 | 74.3 | 74.0 | - | 0.85 |
| CNN + MLP (ECG) [113] | Test set | AS | 10-second | 80.0 | 81.4 | 81.4 | - | 0.88 |
| ValveNet (ECG) [114] | Test set | AS, AR, TR | 10-second | 78.0 | 73.0 | - | - | 0.84 |
| TGNN (PCG) [115] | CV | AS, PS, BAV | Cardiac cycle | 82.8 | 86.7 | 84.2 | - | - |
| Random Forest (PCG) [116] | CV | AS, MR, MS | Cardiac cycle | - | - | 95.1 | - | - |
| SVM (PCG) [119] | CV | AS, MR, MS, AR | Cardiac cycle | - | - | 92.6 | - | - |
| Random Forest (PCG) [117] | CV | AS, MR, MS, MVP | 10-second | 94.8 | 98.7 | 94.8 | - | - |
| SVM (PCG) [120] | CV | AS, MR, MS, MVP | Cardiac cycle | 98.2 | 99.4 | 97.9 | 99.7 | - |
| CNN-BiLSTM (PCG) [121] | CV | AS, MR, MS, MVP | 1.2-second | 98.3 | 99.6 | 99.3 | 98.3 | 0.99 |
| Cardi-Net (PCG) [122] | CV | AS, MR, MS, MVP | 1.4-second | 97.1 | 99.3 | 98.9 | 97.1 | 0.99 |
| Random Forest (PCG) [118] | CV | AS, MR, MS, AR | Cardiac cycle | 97.0 | 99.3 | 98.8 | 97.1 | - |
| Random Forest (SCG/GCG) [71] | CV | AS, AS-MR, AS-MS, AS-TR (coexisting) | 10-second | 97.4 | 97.9 | 96.4 | - | - |
| XGBoost (SCG/GCG) [125] | Test set | AS | 10-second | 97.3 | - | 96.4 | 96.5 | - |
| **MFCC + CMT (ACM)** | **CV** | **AS, AR, MR, PR, TR (coexisting)** | **10-second** | **96.5** | **98.7** | **98.1** | **95.5** | **0.99** |
| **MFCC + CMT (ACM)** | **Test set** | **AS, AR, MR, PR, TR (coexisting)** | **10-second** | **93.2** | **98.0** | **96.8** | **92.9** | **0.98** |

PS: pulmonary stenosis, BAV: bicuspid aortic valve, MVP: mitral valve prolapse, CV: cross-validation was used for performance

when signal morphology changes due to PCG sensor placement [150] and SCG/GCG motion artifacts generated by body movement [151]. X-second signal segmentation on the other hand, is performed based on arbitrarily segmenting the signal into shorter segments, ruling out the classification errors caused by signal annotation. As mentioned in earlier, our signal processing chain is established based on arbitrary 10-second ACM signal segments.

The conventional methods listed in Table 2.11 use a diverse range of sensory data including ECG, PCG, and SCG/GCG for VHD detection. Comparing the performance of our MFCC + CMT (ACM) method with those of DenseNet [112], CNN + MLP [113], and

ValveNet [114] suggests the superiority of our method in terms of sensitivity (93.2% vs. 78.1%, 80.0%, and 78.0% by [112], [113], and [114] respectively) and specificity (98.0% vs. 74.3%, 81.4%, and 73.0% by [112], [113], and [114] respectively). Our framework with an accuracy of 96.8% outperforms the approaches in [112] and [113] with accuracies of 74.0% and 81.4%, respectively. The AUC of 0.98 obtained by our proposed framework demonstrates a superior diagnostic ability compared to CNN-based methods in [112], [113], and [114] with AUCs of 0.85, 0.88, and 0.84 respectively. The substantial difference between the performance of our ACM-driven method and ECG-based approaches can be explained by the components of valve abnormalities which may be missing in the bio-potential profile of the cardiac cycle [152]. Our previous study on the detection of AS with SCG/GCG modalities in [125] achieved a sensitivity, accuracy, and F1 score of 97.3%, 96.4%, and 96.5%, respectively. However, the method was unable to detect other valvular abnormalities.

On the validation set, our sensitivity of 96.5% offers a superior true detection of VHDs compared to TGNN (PCG) and Random Forest (PCG) with sensitivities of 82.8% and 94.8% respectively, and comparable performance to SVM (PCG) [120], CNN-BiLSTM (PCG) [121], Cardi-Net (PCG) [122], Random Forest (PCG) [118], and Random Forest (SCG/GCG) [71] with sensitivities of 98.2%, 98.3%, 97.1%, 97.0%, and 97.4% respectively. This however, is coupled with the fact that our MFCC + CMT (ACM) methodology focuses on the detection of multiple VHD types simultaneously. Our proposed method suggests an excellent performance compared to state-of-the-art methods in terms of specificity (98.7% vs. 86.7%, 98.7%, 99.4%, 99.6%, 99.3%, 99.3%, and 97.9%

by [115], [117], [120], [121], [122], [118], and [71] respectively) and accuracy (98.1% vs. 84.2%, 95.1%, 92.6%, 94.8%, 97.9%, 99.3%, 98.9%, 98.8%, 96.4%, and 96.4% by [115], [116], [119], [117], [120], [121], [122], [118], [71], and [125] respectively). Hence, comparative analyses demonstrate that the proposed MFCC + CMT applied on 10-second ACM recordings can be reliably adopted for the diagnosis of AS, AR, MR, PR, and TR in clinical settings.

## 2.3 Diagnosis of Peripheral Artery Disease Using Backflow Abnormalities in Proximal Recordings of Accelerometer Contact Microphone (ACM)

In this study, we take advantage of the dynamics of blood pressure backflow in obstructed arteries, and establish a novel framework for the detection of PAD. The major contributions of this study are as follows:

- This study introduces a novel method for the diagnosis of PAD using cardiac cycle information collected by a high-precision accelerometer contact microphone. To this end, 95 subjects, including 74 PAD and 21 healthy subjects, were monitored. To the best of our knowledge, this is the first study that addresses the detection of PAD from blood pressure backflow patterns in cardiac cycles.
- A novel feature extraction technique based on transfer learning is introduced that extracts blood pressure backflow patterns and relates them to the presence of PAD and its severity level. The feature extraction approach is based on the use of high-level-of-abstraction feature maps derived from the middle layers of a pre-trained deep neural network.

- A modified vision transformer, coined multi-stream-powered vision transformer, is developed to detect PAD and classify its severity level. This neural network leverages the dynamics of ACM recordings in different frequency bands to recognize patterns associated with peripheral artery disease.

### 2.3.1 Background on Arterial Tree Dynamics

The arterial tree represents the branching system of arteries which are responsible for carrying oxygen-rich blood from the heart to other organs. According to the multi-branch transmission lines (TLs) model introduced by [153], the propagation of blood flow in an artery can be formulated by:

$$q_{out} = \frac{q_{in}^{(1-\Gamma)}}{e^{\gamma l} - \Gamma e^{-\gamma l}}, \qquad (22)$$

where $q_{in}$, $q_{out}$, $\Gamma$, $\gamma$, and $l$ denote the inlet blood flow (mL/s), outlet blood flow (mL/s), reflection coefficient, propagation constant (rad/cm), and arterial length (cm), respectively. For a fully obstructed artery, the left-hand side of (1) holds a zero value since there is no outlet blood flow. Hence, assuming $|q_{in}| > 1$,

$$q_{in}^{(1-\Gamma)} = 0, \qquad (23)$$

which implies either $\Gamma \to +\infty$, or

$$e^{\gamma l} - \Gamma e^{-\gamma l} \to \pm\infty. \qquad (24)$$

Table 2.12 Demographic Information of the Healthy and PAD Participants

| Cohort | Number | Age | BMI (kg/m$^2$) |
|---|---|---|---|
| **Healthy** | 21 (7 M, 14 F) | 64.61 ± 16.43 | 29.15 ± 6.37 |
| **Mild PAD** | 29 (13 M, 16 F) | 71.48 ± 12.05 | 29.20 ± 5.91 |
| **Severe PAD** | 45 (22 M, 23 F) | 75.35 ± 9.08 | 29.24 ± 6.13 |

For a certain artery length of $l$, we can consider $e^{\gamma l}$ and $e^{-\gamma l}$ to be constant positive values. Hence, (3) results in $\Gamma \to \pm\infty$. As the reflection coefficient ($\Gamma$) holds positive values only, the term $\Gamma \to +\infty$ is acceptable. Terms (23) and (24) suggest that the more severe the obstruction, the larger the reflection of the blood flow. The reflection generates a resistive power towards the heart muscle pumping blood into the arterial tree. Consequently, the heart has to pump harder to force the blood through the obstructed arteries, causing the appearance of cardio-mechanical abnormalities in cardiac cycles [154]. In this study, we hypothesize that vibrational patterns associated with backflow in PAD patients emerge in the proximal recordings of

cardiac cycles. To prove our hypothesis, we devise a pattern recognition framework to automatically differentiate healthy subjects from PAD patients based on the vibrational characteristics of their cardiac system.

### 2.3.2 Experimental Setup and Methodology

#### 2.3.2.1 Data Collection and Study Protocol

In this study, seventy-four peripheral artery disease (PAD) patients including 35 males and 39 females were studied at the cardiac care unit of Sorin Medical P.C. The PAD cohort was stratified in terms of severity level into two sub-cohorts of mild-PAD and severe-PAD

with average (± standard deviation) ages of 71.48 (± 12.05) and 75.35 (± 9.08) years, respectively. Additionally, a control group consisting of 7 healthy males and 14 healthy females with average (± standard deviation) age of 64.61 (± 16.43) years were measured for comparison with the PAD cohort. Further demographic information on the subjects is summarized in Table 2.12.

Fig. 2.18 illustrates the proposed framework for the detection of PAD. As shown in Fig. 2.18 (a), the experimental setup consists of a $\pm 4$ g sensitive accelerometer contact microphone (ACM) with micro-g resolution [128,129] (obtained from StethX Microsystems Inc., Atlanta, USA) attached to the chest wall along the third rib using medical-grade adhesive tape. Additionally, the experimental setup includes a three-lead ECG sensor node (ECG Development Kit; Shimmer Sensing, Dublin, Ireland) with the electrodes attached to the right arm (RA), left arm (LA), right leg (RL), and left leg (LL),

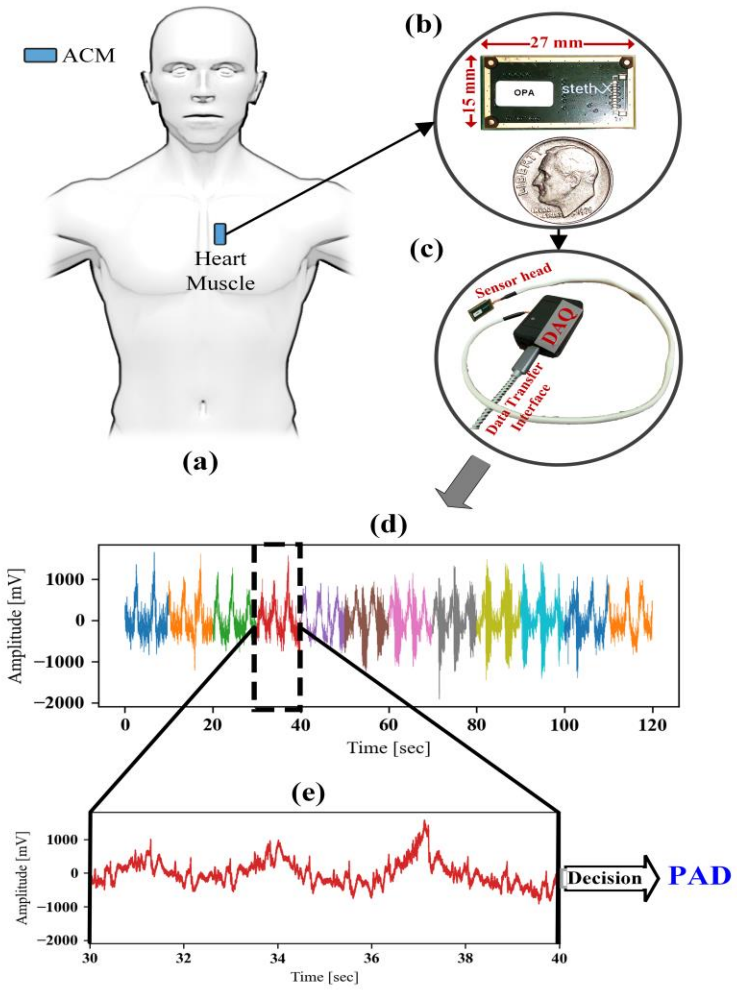


Fig. 2.18 The proposed PAD detection framework. (a) The sensor arrangement on the torso, (b) dimensions of the accelerometer contact microphone, (c) the data acquisition module, (d) the ACM signal segmentation, and (e) decision making based on a signal segment.

as well as a photoplethysmography (PPG) sensor (Shimmer3 GSR; Shimmer Sensing, Dublin, Ireland) attached to the ear-lobe using an ear clip. The ACM sensor head has a small form factor of 27 mm × 15 mm × 2.5 mm which is comparable to the size of a dime as depicted in Fig. 2.18 (b). This device is a low-noise accelerometer with a wide operational bandwidth of 0-10 kHz, allowing for recording heartbeat-induced sounds and

vibrations on the chest wall. This device is not sensitive to airborne emission sounds, making it a robust phonocardiogram sensor against acoustic ambient noise. The data is digitized using a 24-bit analog-to-digital converter integrated in the sensor head, and transferred to a computer in real-time through a data transfer cable connected to the data acquisition (DAQ) module, as shown in Fig. 2.18 (c). The subjects were seated at rest on a chair for a period of five minutes, followed by five minutes of ACM measurements at a sampling rate of 22.33 kHz, still at a seated position. The patient experimental protocol was approved by the Institutional Review Board of Stevens Institute of Technology under protocol number 2022-044 (N). It is to be noted that ECG and PPG data are not used in this specific study. In the proposed framework, the objective is to detect peripheral artery disease using short segments of ACM recordings as shown in Figs 2.18 (d) and (e).

#### 2.3.2.2 Data Preparation

In order to reduce the computational complexity of the processing chain, ACM recordings were initially down-sampled to a sampling rate of 4 kHz. This practice did not affect the heart sound components which fall below 1 kHz [130]. The ACM signals were then segmented into 10-second overlapping windows with 90% overlap between consecutive windows. The choice of the window length follows the superior results achieved for the 10-second duration compared to other segment lengths in [125]. Depending on the sensor placement and the body mass index (BMI) of the subject, the signal-to-noise ratios (SNRs) of the recorded signals vary within the range of [5, 15] dB. This motivates the need for a denoising algorithm to enhance the quality of heart sounds.

Fig. 2.19 depicts two examples of ACM raw recordings with SNR values of 10.5 dB and 5.3 dB. The denoising algorithm begins with zero-padding the beginning and end of each segment by 256 points (62.5 ms) to prevent the transient effect of filtering conducted in the following processing steps. As shown in Fig. 2.19 (d), heart sound components of the low-SNR phonocardiogram signal are represented by small peaks modulated on the low-frequency waveforms corresponding to respiration. A differentiation filter was applied to the signal segment to amplify high-frequency components. Each segment was then filtered by a 3rd zero-phase Butterworth filter with a cut-off frequency of 10 Hz to remove respiratory components and baseline wander. A cumulative filter was used to reconstruct the heart sound morphology, followed by discarding the padded points at the beginning and end of each segment. Figs 2.19 (b) and (e) respectively illustrate the filtered ACM segments with SNR 10.5 dB and 5.3 dB. According to Fig. 2.19 (e), the remaining signal is still contaminated with high-energy noise, hindering the full appearance of the first and second heart sounds which are known as S1 and S2, respectively.

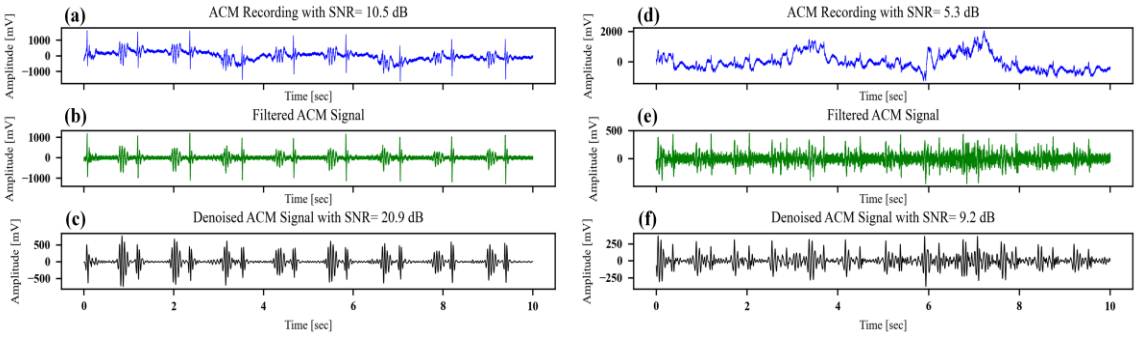


Fig. 2.19 Signal denoising examples for high-SNR and low-SNR ACM segments. (a) High-SNR ACM raw recording (SNR= 10.5 dB), (b) filtered high-SNR ACM signal segment, (c) denoised high-SNR signal segment generated by EMD, (d) low-SNR ACM raw recording (SNR= 5.3 dB), (e) filtered low-SNR ACM signal segment, and (f) denoised low-SNR signal segment generated by EMD.

Several studies have demonstrated the potentials of empirical mode decomposition (EMD) for phonocardiogram denoising [131,155]. The core idea behind EMD is to adaptively decompose an oscillatory signal into a series of zero-mean amplitude- and frequency-modulated signals called intrinsic mode functions (IMFs) through an iterative procedure. EMD begins with interpolating between local maxima and local minima to obtain upper and lower envelopes, respectively. These two envelopes are then averaged, and the resulting component is subtracted from the original signal. The resulting signal is called an IMF only if it holds a zero mean, and the number of local extrema and zero crossings are equal or differ by at most 1. The IMF is then subtracted from the original signal to produce the residue signal which plays the role of the original signal for the next iteration. This process repeats until the final residue signal is a constant or monotonic function. Each filtered ACM signal segment ($x(n)$) undergoes EMD which results in:

$$x(n) = \sum_{i=1}^{L-1} h^{\mathrm{i}}(n) + r^{L}(n), \qquad (25)$$

where $h^{\mathrm{i}}(n)$, $L-1$, and $r^{L}(n)$ denote the $i^{th}$ IMF, number of IMFs, and the final residue signal respectively. In this work, we propose to denoise the signal segment by selecting the most relevant IMFs. The selection criteria can be defined either in the frequency domain [44] or time domain [131], where the latter has shown promising results for heart sound denoising. Hence, a modified version of the method presented in [131] is designed to incorporate only relevant and smooth IMFs for SNR enhancement. To this end, the actual energy of each IMF is compared with the estimated energy ($V_i$) proposed by [131], as mentioned below:

$$E_i = \frac{E_1}{\beta}\alpha^{-i},\ i = 2, 3, 4, \dots,\ L, \qquad (26)$$

where $\alpha$ and $\beta$ denote the estimation parameters which are 2.01 and 0.719 respectively. $E_1$ represents the estimated energy for the first IMF which is calculated using the median absolute deviation (MAD) as follows:

$$E_1 = \left(\frac{\text{med}\{|h^1 - \text{med}\{h^1\}|\}}{\rho}\right)^2, \qquad (27)$$

with $\text{med}\{.\}$ and $\rho$ being the median operator and 0.6745 according to [155], respectively. Each IMF whose actual energy exceeds its estimated energy is included for signal reconstruction. In the original denoising method, relevant IMFs are then further denoised using a thresholding-based method to remove suspect noise components. In our signal processing chain however, the desire is to preserve potential patterns corresponding to PAD. To this end, we replace the thresholding technique with a 3rd-order Savitzky-Golay smoothing filter of size 50 ms to produce smooth outputs while maintaining signal patterns [133]. As illustrated in Figs 2.19 (c) and (f), the denoising algorithm has enhanced the signal quality by 10.4 dB and 3.9 dB for high-SNR and low-SNR signal segments respectively.

#### 2.3.2.3 Feature Extraction

Blood backflow associated with obstructions in PAD patients is expected to manifest abnormal patterns in the energy of ACM recordings and their corresponding dynamics. Inspired by cardiologists who use stethoscopes to inspect the rhythm and intensity of the heart sounds, we extract energy-related features to automatically diagnose PAD. These features are summarized as follows:

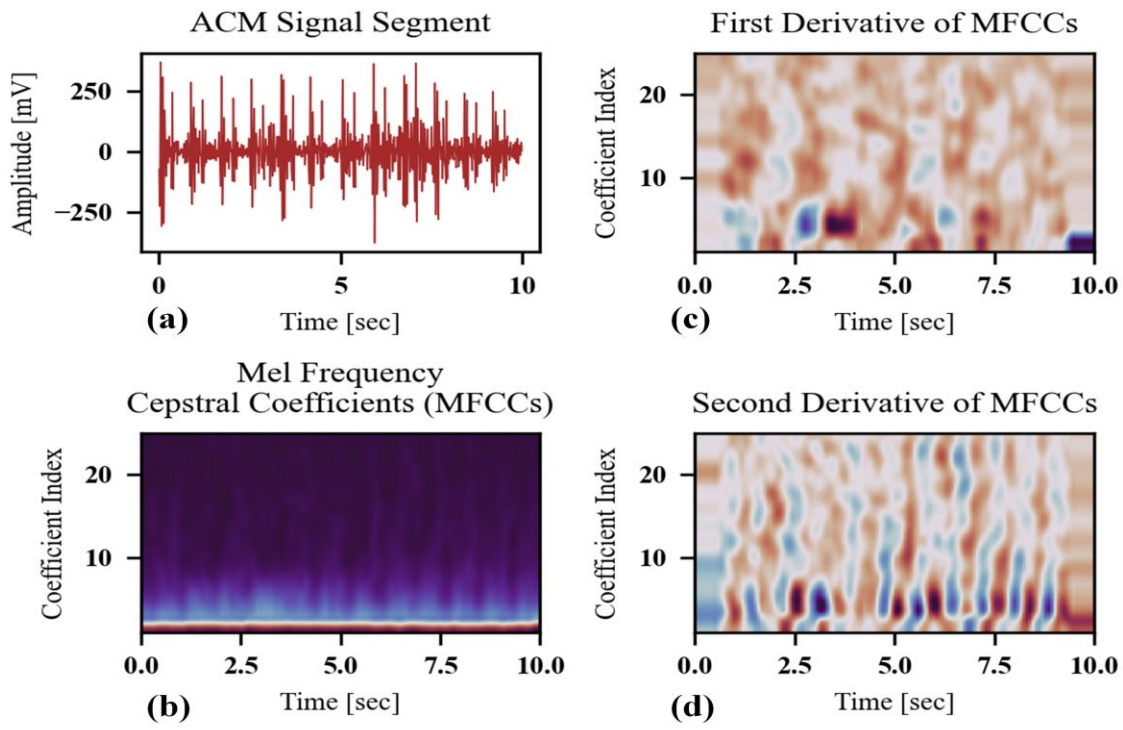


Fig. 2.20 Heart sound feature extraction. (a) denoised ACM signal segment, (b) Mel frequency cepstral coefficients (MFCCs), (c) the first derivative of MFCCs, and (d) the second derivative of MFCCs.

#### 2.3.2.3.1 Mel frequency cepstral coefficients (MFCCs)

These coefficients distribute the short-term power spectrum of a signal in accordance with the human auditory system perception in a scale called Mel [134]. This scale uses a filter spaced linearly at frequencies below 1 kHz and has logarithmic spacing above 1 kHz. MFCC computation begins with framing a signal segment into shorter frames of 0.5 seconds (2,048 samples) with 25% overlap (512 samples) between consecutive frames. For each frame, the periodogram is then calculated and filtered through a Mel filter bank with 25 triangular filters. The energy of the signal is summed per filter, thus generating 25 coefficients. Finally, the discrete cosine transform is applied to the coefficients to obtain MFCCs. Hence, for each 10-second signal segment hence, we would

have a 2-D representation of size 25 × 81, representing 81 frames, each involving 25 coefficients. Figs 2.20 (a) and (b) depict a signal segment and its corresponding MFCCs respectively. In order to model the dynamics of the heart sounds, the first and second derivatives of MFCCs are calculated by subtracting the coefficients of successive time frames once for the first derivative and twice for the second derivative as shown in Figs 2.20 (c) and (d), respectively.

#### 2.3.2.3.2 High-level-of-abstraction coefficients (HLACs)

Convolutional neural networks (CNNs) have shown excellent promise for image-related tasks such as image classification [156], segmentation [157], and denoising [158]. In addition, they are also capable of extracting discriminative features from raw data at

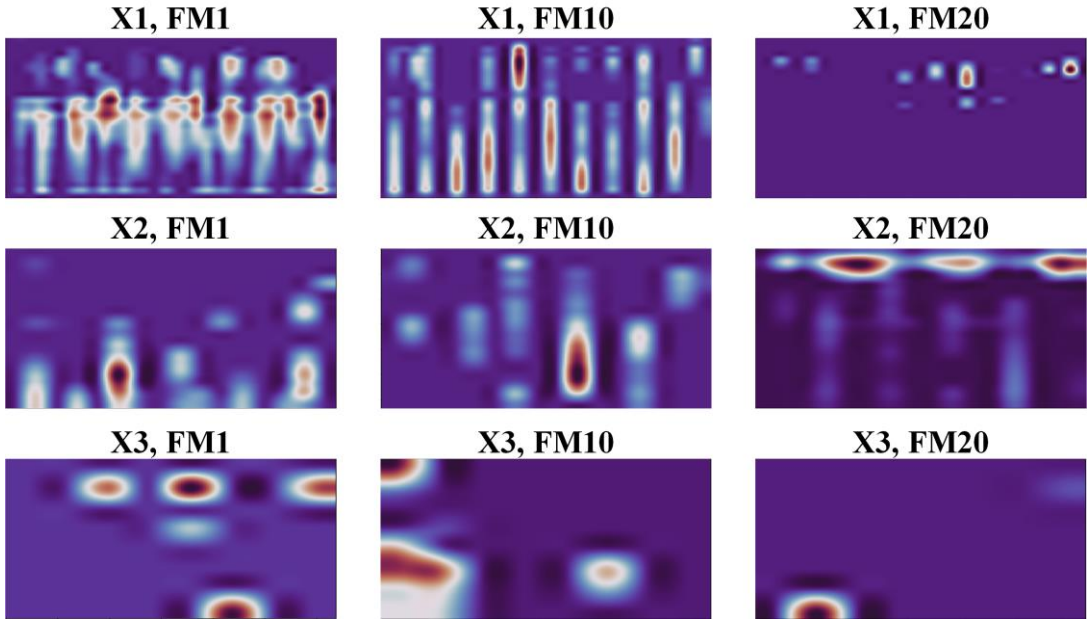


Fig. 2.21 Examples of high-level-of-abstraction coefficients extracted from MFCCs and its derivatives by ResNet50. Top, middle, and bottom rows correspond to feature maps (FM1, FM10, FM20) from layers 10, 22, and 40 respectively.

different levels of abstraction [159]. We use this capability of CNNs to extract features from MFCCs which can be considered as an input image to a 2-D CNN architecture. To this end, we adopt a pre-trained residual network with 50 layers (ResNet50) [160], and fine-tune the network on our own dataset. This practice, which is called transfer learning, allows us to reuse RseNet50 pre-trained on the ImageNet dataset for the PAD detection task. Leveraging residual layers in ResNet prevents vanishing gradient issues which generally occur in deep networks. As such, ResNet50 is capable of providing features with a variety of abstraction levels. In this work, we use the outputs of layers 10, 22, and 40 of a ResNet50 which have 256, 512, and 1,024 feature maps (FMs), respectively. Fig. 2.21 depicts a few examples of feature maps 1, 10, and 20, corresponding to layers 10, 22, and 40 which are represented by X1, X2, and X3 respectively. As can be observed from top to bottom in Fig. 2.21, transfer learning-based HLACs provide both fine and coarse features of the input, helping with decision-making about the presence of PAD. As a result, three feature maps associated with MFCCs and their first and second derivatives, denoted by X, are combined with HLACs to contribute to the accurate detection of PAD.

#### 2.3.2.4 Multi-Stream-Powered Context-Aware Prediction

As mentioned earlier, the cardiac muscle should pump the blood harder to cope with backflow pressure waves caused by PAD. Abnormal changes in pressure and thus energy level of the ACM recordings are the keys to PAD detection, for which MFCCs and HLACs are derived. The abnormality associated with PAD may appear in any sub-band or time point. This motivates the need for a 2-D context-aware predictive model to characterize inter-dependencies within various sections of a feature map.

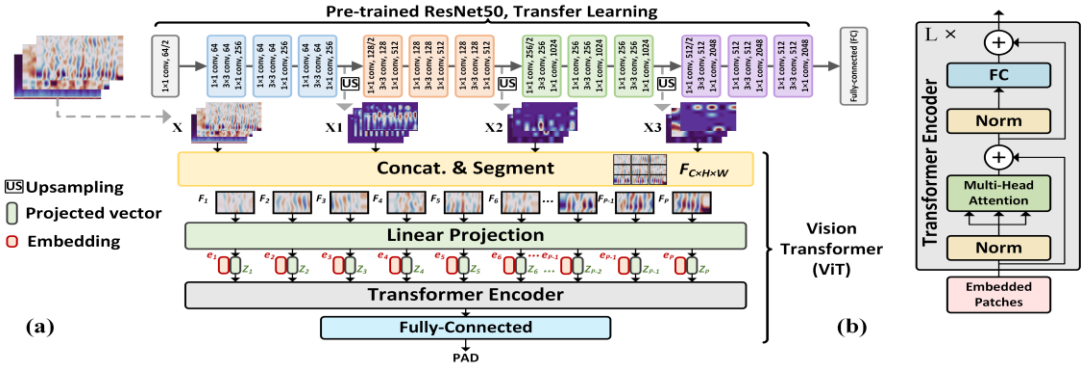


Fig. 2.22 Multi-stream-powered vision transformer (MSPViT) architecture for PAD detection network. (a) a ResNet50 for HLACs extraction followed by a vision transformer (ViT) for PAD detection, and (b) a transformer encoder which employs a multi-head attention module for modeling inter-patch dependencies.

Transformers are widely used context-aware models in natural language processing [143], which outperform their ancestors such as long short-term memory (LSTM) networks which were formerly adopted for machine translation tasks and time series modeling. The advantage of transformers over conventional methods is that sequences are modeled as a whole using self-attention modules, whereas LSTM models a sequence on an element-wise basis. In our application however, we cannot use the original structure of the transformer since input features i.e., MFCCs and HLACs, are two-dimensional features. Vision transformer (ViT) is a recently proposed transformer which is customized for image classification [142]. In this work, we develop a network based on ViT which is powered by a multi-stream ResNet50 for PAD detection, as illustrated in Fig. 2.22.

As shown in Fig. 2.22 (a), a pre-trained ResNet50 on the 1,000-class ImageNet dataset is fine-tuned on the PAD dataset. For this task, the fully-connected (FC) layer at the end of the original ResNet50 with a length of 1,000 is replaced with a fully-connected layer with either two neurons for PAD detection, or three neurons for severity classification. As mentioned earlier, three MFCCs-related feature maps (X) and the outputs

of layers 10 (X1), 22 (X2), and 40 (X3) which are HLACs feature maps, are concatenated to build the input feature maps to the vision transformer network. As mentioned earlier, the dimension of MFCCs is 25 × 81. We resize MFCCs and the corresponding first and second derivatives through a bilinear interpolation to 128 × 128 to have standard-size inputs. This results in X1, X2, and X3 being of various sizes of 32 × 32, 16 × 16, and 8 × 8, respectively. All feature maps X1, X2, and X3 should therefore be reshaped to 128 × 128 using bilinear interpolation, so these features would be concatenable with the resized MFCCs-related features. As a result, the concatenation builds a larger set of feature maps consisting of 1,795 (3 + 256 + 512 + 1,024) feature maps of size 128 × 128.

As shown in Fig. 2.22 (a), concatenated feature maps are denoted by $F \in \mathbb{R}^{C \times H \times W}$ with $C$, $H$, and $W$ representing the number of channels (1,795), height (128), and width (128) of the feature maps, respectively. In the vision transformer, $F$ is segmented into $P$ small patches represented by $F_i \in \mathbb{R}^{C \times p \times p}, i = 1, 2, 3, \dots, P$ , where $p$ shows the dimensions of the patches. Each $p \times p$ patch is then reshaped to a vector with a length $C \times p \times p$. This practice allows for processing images as 1-D vectors, similar to the processing of time series using a transformer network. In this work, the size of each patch is set to $p = 8$, thus resulting in 256 patches for an input feature size of 128 × 128. A linear projection layer is then applied to each 114,800-long flattened vector to shrink it to a vector of length 64. As a result, $P$ vectors of size 32 ($Z_i \in \mathbb{R}^{1 \times 64}, i = 1, 2, 3, \dots, P$) are generated, and summed with trainable embedding vectors which are denoted by $e_i \in \mathbb{R}^{1 \times 64}, i = 1, 2, 3, \dots, P$ in Fig. 2.22 (a). The embeddings enable the network to take the position of each patch in the image into account while learning the patterns of healthy and PAD subjects.

To encode the dependencies among the patches of a whole sequence, we employ $L = 4$ stacked transformer encoders, the architecture of each is illustrated in Fig. 2.22 (b). This encoder consists of two normalization layers, 8 multi-head attention layers, and a fully-connected (FC) as well as two residual layers. Normalization layers accelerate the learning process and enhance the generalizability of the model [161]. After the batch normalization step, a multi-head attention module is employed to quantify the dependencies among the embeddings corresponding to patches. The attention mechanism begins with multiplying each embedding with three learnable matrices $W^Q \in \mathbb{R}^{32\times64}$, $W^K \in \mathbb{R}^{32\times64}$, and $W^V \in \mathbb{R}^{32\times64}$. This generates three vectors called query ($q_i \in \mathbb{R}^{1\times32}$), key ($k_i \in \mathbb{R}^{1\times32}$), and value ($v_i \in \mathbb{R}^{1\times32}$) for each embedding. The query of each embedding is multiplied by all the keys in the sequence to find the relevance between the embedding under test and other elements in the sequence. The score vector which is of size $P$ subsequently undergoes a Softmax function to rescale the scores into the range of [0, 1], and make them sum up to 1. The resulting rescaled scores are then multiplied by their corresponding value vectors to build the attention module output for the embedding under test. Hence, the attention module can be formulated as:

$$U = Softmax\left(\frac{\mathrm{Q}\times\mathrm{K}^{\mathrm{T}}}{\sqrt{32}}\right)V, \qquad (28)$$

where $Q \in \mathbb{R}^{P\times32}$, $K \in \mathbb{R}^{P\times32}$, and $V \in \mathbb{R}^{P\times32}$ denote the matrices built by stacking query vectors, key vectors, and value vectors respectively. Matrix $U \in \mathbb{R}^{P\times32}$shows the output matrix with $P$ rows, each of which represents the output of the attention layer corresponding to a patch (or embedding). In our implementation, we embed 8 attention

heads for each attention layer, yielding 8 attention outputs $U_i, i = 1, 2, 3, \dots, 8$. Per embedding then, a vector of size $8 \times 32$ results. Every output is finally converted to a vector of size 256 through a fully-connected layer in the transformer encoder as shown in Fig. 2.22 (b). Another fully connected layer is employed in the PAD detection network in Fig. 2.22 (a) to resize the encoder output to the number of classes, i.e., two for PAD identification and three for PAD severity level classification. The objective of the multi-stream-powered vision transformer (MSPViT) is to minimize the error between predicted values and the corresponding ground truth labels, using mean squared error (MSE) as follows:

$$\ell(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2 , \qquad (29)$$

where $y$ and $\hat{y}$ denote one-hot coding vectors for the ground truth and predicted values, respectively.

### 2.3.3 Experimental Results

This section provides an overview of the training procedure and the metrics used to evaluate the performance of MSPViT. Additionally, the performance of the model on PAD detection and severity classification is discussed in the following sub-sections.

#### 2.3.3.1 Training Procedure

As mentioned earlier, each signal segment is represented by its MFCCs and HLACs which are used as the input to the MSPViT. This network is trained in a supervised manner to learn the subjects' labels assigned by two interventional cardiologists based on echocardiography parameters. In this work, we trained the model for both the detection of

PAD (binary classification) and the classification of PAD severity levels into either healthy, mild PAD, or severe PAD (ternary classification).

The training procedure was conducted using the data of 95 subjects amounting to 46,878 signal segments. A 10-fold leave-subject-out cross-validation (10-LSOCV) approach was adopted to assess the generalizability of the model to unseen data. To this end, the dataset was split into train (68 subjects; 70%), validation (18 subjects; 20%), and test (9 subjects; 10%) sets. The model is trained on the train set for 150 epochs, validated on the validation set, and the performance is reported using the test set when the loss function, defined in (8), is minimized. The Adam optimization algorithm [147] was employed with an initial learning rate of 0.001 which was reduced by a factor of 0.98 every ten epochs without performance improvement. The model was implemented on a 24-GB NVIDIA GeForce RTX 3090 Ti FTW3 with a batch size of 64.

#### 2.3.3.2 Evaluation Metrics

Statistical analyses were conducted to evaluate the performance of MSPViT and compare it with state-of-the-art PAD detection methods. In order to compare the ground-truth labels with those predicted by MSPViT, sensitivity (Sen), specificity (Spec), accuracy (Acc), positive predictive value (PPV), and F1 score (F1) are calculated as follows:

$$\text{Sen} = 100 \times \frac{TP}{TP+FN}, \qquad (30)$$

$$\text{Spec} = 100 \times \frac{TN}{TN+FP}, \qquad (31)$$

$$\text{Acc} = 100 \times \frac{TP}{TP+FP+FN}, \qquad (32)$$

Table 2.13 10-LSOCV Average (± Standard Deviation) Performance on PAD Detection

| Architecture | Sen (%) | Spec (%) | Acc (%) | PPV (%) | F1 (%) |
|---|---|---|---|---|---|
| **ResNet50 (X)** | 92.69 (± 4.51) | 90.27 (± 11.44) | 91.55 (± 4.99) | 96.04 (± 4.48) | 94.24 (± 3.40) |
| **ViT (X)** | 99.08 (± 0.71) | 93.36 (± 5.48) | 97.52 (± 1.87) | 97.61 (± 2.66) | 98.32 (± 1.36) |
| **MSPViT (X1)** | 99.05 (± 0.84) | 98.96 (± 1.0) | 99.05 (± 0.67) | 99.68 (± 0.31) | 99.36 (± 0.47) |
| **MSPViT (X2)** | **99.45 (± 0.47)** | **98.21 (± 2.93)** | **99.07 (± 0.89)** | **99.30 (± 1.25)** | **99.37 (± 0.63)** |
| **MSPViT (X3)** | 98.92 (± 0.97) | 93.47 (± 7.74) | 97.20 (± 2.50) | 97.31 (± 3.58) | 98.07 (± 1.79) |
| **MSPViT (X, X1, X2, X3)** | 98.66 (± 1.02) | 96.85 (± 2.76) | 98.04 (± 0.97) | 98.68 (± 1.53) | 98.65 (± 0.73) |

$$\text{PPV} = 100 \times \frac{TP}{TP+FP}, \qquad (33)$$

$$\text{F1} = \frac{2\times\text{PPV}\times\text{Sen}}{\text{PPV}+\text{Sen}}, \qquad (34)$$

where $TP$, $TN$, $FP$, and $FN$ denote the true positive (correctly detected as PAD), true negative (correctly detected as healthy), false positive (falsely detected as PAD), and false negative (falsely detected as healthy) cases, respectively. It is to be noted that statistical metrics were calculated for each label in the ternary classification, and each metric was averaged over all three classes. i.e., macro averaging. Sensitivity and specificity respectively indicate the concordance of MSPViT with respect to the reference, whereas the positive predictive value signifies the likelihood that MSPViT can successfully identify PAD. Additionally, the F1 score was adopted to present a harmonic mean of positive predictive value and sensitivity of MSPViT.

#### 2.3.3.3 PAD Detection Performance

Table 2.13 summarizes the results of 10-LSOCV for PAD detection (binary classification). In this table, MSPViT (X1), MSPViT (X2), and MSPViT (X3) respectively

Table 2.14 10-LSOCV Average (± Standard Deviation) Performance on PAD Severity Classification

| Architecture | Sen (%) | Spec (%) | Acc (%) | PPV (%) | F1 (%) |
|---|---|---|---|---|---|
| **ResNet50 (X)** | 95.27 (± 5.15) | 96.89 (± 2.46) | 97.25 (± 2.11) | 96.07 (± 4.04) | 95.58 (± 4.72) |
| **ViT (X)** | 89.16 (± 6.53) | 94.30 (± 3.61) | 95.38 (± 3.34) | 94.66 (± 2.98) | 91.12 (± 5.09) |
| **MSPViT (X1)** | 96.31 (± 1.59) | 95.94 (± 1.58) | 96.22 (± 1.53) | 93.72 (±2.79) | 94.70 (± 1.97) |
| **MSPViT (X2)** | 97.60 (± 1.12) | 96.89 (± 1.36) | 96.97 (± 1.32) | 93.61 (± 3.34) | 95.18 (± 2.06) |
| **MSPViT (X3)** | **96.66 (± 4.08)** | **97.34 (± 2.19)** | **97.65 (± 1.67)** | **96.24 (± 2.51)** | **96.29 (± 3.01)** |
| **MSPViT (X, X1, X2, X3)** | 96.26 (± 1.42) | 96.92 (± 1.19) | 97.16 (± 1.13) | 95.28 (± 2.68) | 95.66 (± 1.92) |

denote vision transformers supplied with the feature maps of layers 10, 22, and 40 of a ResNet50 fine-tuned on the PAD data. The last row signifies the results associated with MSPViT (X, X1, X2, X3) whose input incorporates MFCCs and their derivatives (three feature maps) as well as the feature maps corresponding to layers 10, 22, and 40 of a pre-trained ResNet50, amounting to a total of 1,795 feature maps. Additionally, the performance of a ResNet50-alone and a vision transformer, denoted by ResNet50 (X) and ViT (X) respectively, are listed in the table for the sake of comparison.

As shown in the table, MSPViT (X2) with 99.45% (± 0.47) offers the highest sensitivity among other architectures for PAD detection, implying the lowest missed rate of PAD cases. ViT (X) and MSPViT (X1) respectively with 99.08% (± 0.71) and 99.05% (± 0.84) of sensitivity scores are the runner ups for PAD detection. MSPViT (X1) and MSPViT (X2) identify PAD with specificity scores of 98.96% (± 1.0) and 98.21% (± 2.93) respectively, which are superior to those of other architectures. In parallel with specificity, MSPViT (X1) with 99.68% (± 0.31) and MSPViT (X2) with 99.37% (± 0.63) have the highest PPV, thus are most likely to successfully detect PAD. According to the reported

F1 scores, vision transformers powered by high-level-of-abstraction features from layers 10 and 22, i.e., MSPViT (X1) and MSPViT (X2), predict the presence of PAD with 99.36% (± 0.47) and 99.37% (± 0.63) respectively. The highest F1 score, i.e., 99.37%, implies increases of 5.13% and 1.05% compared to the ResNet50-alone and ViT-alone architectures, respectively. MSPViT (X1) and MSPViT (X2) therefore, suggest the best performance for PAD detection.

#### 2.3.3.4 Severity Classification Performance

Statistical analyses on PAD severity classification are presented in Table 2.14. Similar to PAD detection, we quantify the performance of MSPViT in comparison to ResNet50 (X) and ViT (X) using sensitivity, specificity, accuracy, PPV, and F1 scores. As summarized in Table 2.14, the highest sensitivity scores were achieved by MSPViT (X2) and MSPViT (X3) with 97.60% (± 1.12) and 96.66% (± 4.08) respectively. An average specificity of 97.37% (± 2.19) ranks the MSPViT (X3) architecture at the top for false stratification of PAD severity. MSPViT (X3) and MSPViT (X, X1, X2, X3) suggest the highest average PPV scores of 96.24% (± 2.51) and 95.28% (± 2.68), respectively, presenting superior performance compared to other architectures with 96.07% (± 4.04), 94.66 (± 2.98), 93.72% (± 2.79), and 93.61% (± 3.34) for ResNet50 (X), ViT (X), MSPViT (X1), and MSPViT (X2) respectively. In terms of F1 score, MSPViT with 96.29% (± 3.01) outperforms MSPViT (X, X1, X2, X3), ResNet50 (X), ViT (X), MSPViT (X1), and MSPViT (X2) with 95.66% (± 1.92), 95.58% (± 4.72), 91.12% (± 5.09), 94.70% (± 1.97), and 95.18% (± 2.06) respectively, which demonstrates the superiority of MSPViT (X3) for the classification of PAD severity levels.

Comparing Tables 2.13 and 2.14 signifies the importance of HLACs features as well as the functionality of the layers for PAD detection and severity classification. While layer 22 (lower abstraction) provides a discriminative feature space for PAD detection, i.e., binary classification, severity level classification could benefit from higher abstraction level features in layer 44. This behavior can be related to the ternary classification defined for PAD severity stratification which requires more detailed information in comparison to PAD detection.

#### 2.3.3.5 Receiver Operating Characteristic (ROC)

The discrimination power of MSPViT is also evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC). In this work, ROC is defined as the true positive rate versus the false positive rate for each class as illustrated in Fig. 2.23. The predicted labels of the test data from all folds were appended, and the ROC curve was obtained using all the test data. Fig. 2.23 (a) indicates the ROC curve associated with PAD detection (binary classification), whereas Figs 2.23 (b), (c), and (d) correspond to the ROC curves of healthy, mild-PAD, and severe-PAD classes respectively. The highest AUC for PAD detection was achieved by MSPViT (X1) with 0.997, followed by MSPViT (X3) and MSPViT (X2) with AUC of 0.994 and 0.990 respectively. ResNet50 (X) offers the lowest AUC of 0.923 which still implies excellent prediction stability. The performance obtained in Fig. 2.23 (a) confirms the validity of the assumption of backflow pressure resulting from arterial obstruction associated with PAD in the arterial tree, mentioned earlier.

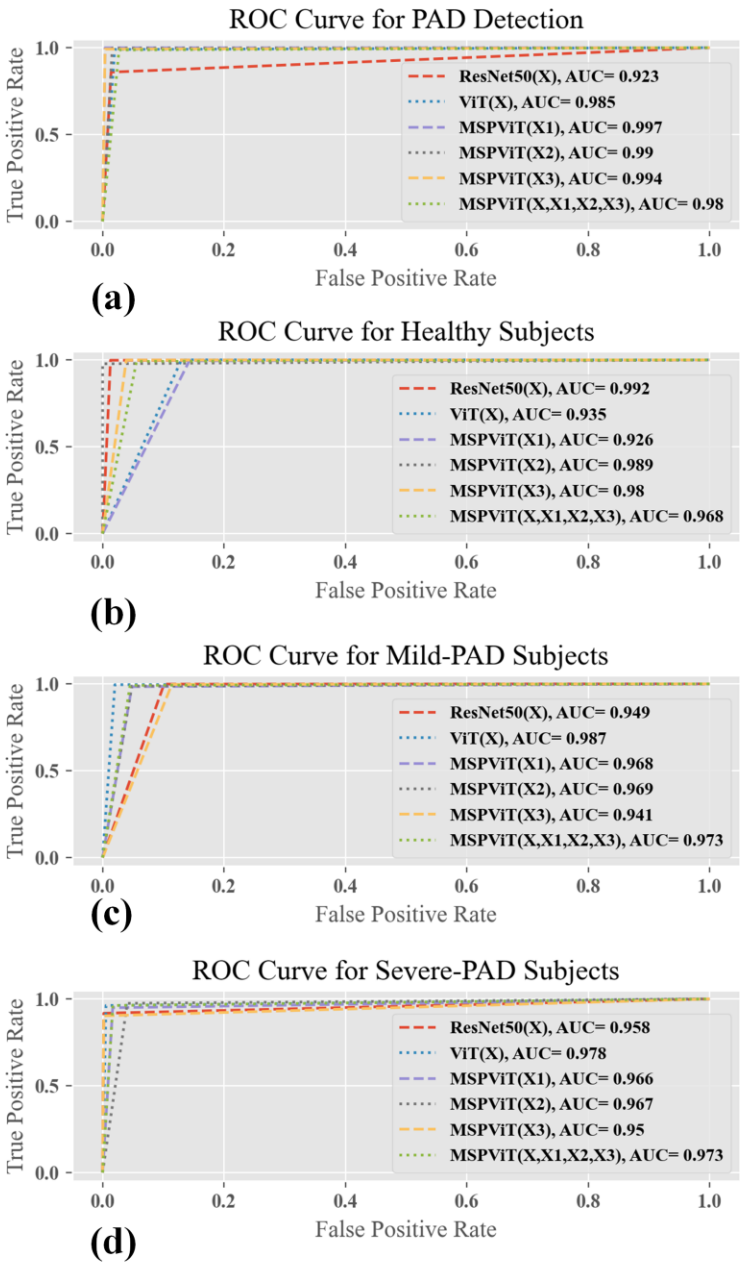


Fig. 2.23 Receiver operating characteristic (ROC) curves for (a) PAD binary classification, (b) healthy class in ternary classification, (c) mild-PAD class in ternary classification, and (d) severe-PAD class in ternary classification.

In order to present ROC characteristics in PAD severity classification, we compare each class with others in a one-versus-rest manner. For this purpose, the target class was labeled

with one, and the other two classes were assigned zero. As such, three ROC curves associated with healthy, mild-PAD, and severe-PAD classes were achieved as shown in Figs 2.23 (b), (c), and (d) with AUCs within the ranges of [0.926, 0.992], [0.941, 0.987], and [0.950, 0.978], respectively. Among the MSPViT architectures, the highest AUC scores for healthy, mild-PAD, and severe-PAD cases were achieved by MSPViT (X2) with AUC = 0.989, MSPViT (X2) with AUC = 0.969, MSPViT (X, X1, X2, X3) with AUC = 0.973 respectively, although MSPViT (X3) with $\text{AUC}_{healthy} = 0.980$, $\text{AUC}_{mild} = 0.941$, and $\text{AUC}_{severe} = 0.950$ suggests slight differences compared to their relative superior models. Hence, MSPViT (X3) predicts the severity level with an average AUC of 0.957.

### 2.3.4 Discussion

In the previous section, we comprehensively investigated the feasibility of the detection of PAD and classification of the severity levels through attention-based modules supplied by MFCCs and HLACs. In this section, the clusters corresponding to different classes in the output of the MSPViT are measured to analyze the separability of classes in a high-dimensional feature space. Additionally, our framework for PAD detection and classification is compared with state-of-the-art methods. Finally, the limitations of the methods are comprehensively discussed.

#### 2.3.4.1 High-Dimensional Separability

A well-trained deep neural network classifier is expected to discriminate test samples of different classes in the high-dimensional space presented by the last layer of the network. We use the output of the transformer encoder layer to inspect separability among PAD classes. Principal component analysis (PCA) was utilized for mapping 32-

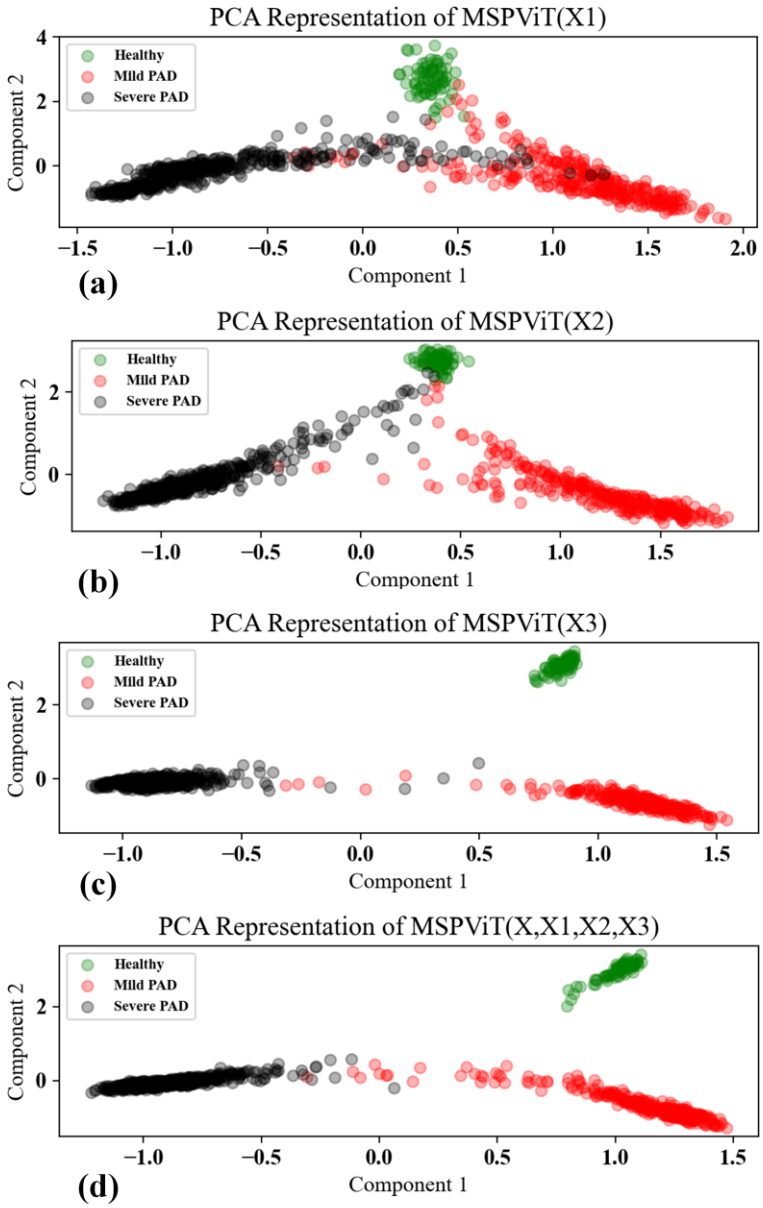


Fig. 2.24 Principal component analysis (PCA) applied to the outputs of the transformer encoder corresponding to (a) MSPViT (X1), (b) MSPViT (X2), (c) MSPViT (X3), and (d) MSPViT (X, X1, X2, X3).

dimensional feature vectors into a 2-dimensional space as shown in Fig. 2.24 to visually inspect the separability for fold 1 of test data. As observed in Figs 2.24 (a), (b), (c), and (d) which respectively correspond to architectures MSPViT (X1), MSPViT (X2), MSPViT

Table 2.15 Silhouette Scores of Obtained Clusters for PAD Detection Network in 10-LSOCV

| Fold | MSPViT (X1) | MSPViT (X2) | MSPViT (X3) | MSPViT (X, X1, X2, X3) |
|---|---|---|---|---|
| **1** | 0.75 | 0.82 | 0.76 | 0.75 |
| **2** | 0.40 | 0.55 | 0.49 | 0.42 |
| **3** | 0.76 | 0.83 | 0.82 | 0.79 |
| **4** | 0.77 | 0.83 | 0.79 | 0.80 |
| **5** | 0.48 | 0.57 | 0.63 | 0.62 |
| **6** | 0.78 | 0.77 | 0.89 | 0.85 |
| **7** | 0.45 | 0.60 | 0.41 | 0.42 |
| **8** | 0.39 | 0.41 | 0.45 | 0..37 |
| **9** | 0.41 | 0.44 | 0.38 | 0.44 |
| **10** | 0.75 | 0.77 | 0.84 | 0.75 |
| **Average (± Std.)** | **0.59 (± 0.18)** | **0.66 (± 0.16)** | **0.65 (± 0.20)** | **0.65 (± 0.18)** |

(X3), and MSPViT (X, X1, X2, X3), excellent separability is offered among healthy, mild-PAD, and severe-PAD classes. Slight overlaps between classes are due to PCA discarding non-principal components which potentially provide discriminating capabilities. As such, we employ the whole feature vectors which are of length 32, and quantify the clusters formed by each class in the 32-dimensional space. To this end, we adopt the silhouette score for each test data point which is defined as follows:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}, \qquad (35)$$

where $b(i)$ and $a(i)$ denote the distance between a sample and the nearest cluster that the sample does not belong to and the distance between a sample and the cluster that the sample is a part of, respectively. Silhouette score ranges within [-1, +1] with +1 suggesting a perfect match between the sample and the assigned label, i.e., healthy, mild-PAD, and severe-PAD. In this work, we present the silhouette score by averaging over the test data points corresponding to 10-LSOCV. Tables 2.15 and 2.16 summarize silhouette scores for folds of PAD detection and severity classification, respectively. Table 2.15 suggests the

Table 2.16 Silhouette Scores of Obtained Clusters for Severity Classification Network in 10-LSOCV

| Fold | MSPViT (X1) | MSPViT (X2) | MSPViT (X3) | MSPViT (X, X1, X2, X3) |
|---|---|---|---|---|
| **1** | 0.63 | 0.66 | 0.72 | 0.76 |
| **2** | 0.66 | 0.74 | 0.77 | 0.73 |
| **3** | 0.56 | 0.66 | 0.87 | 0.83 |
| **4** | 0.59 | 0.67 | 0.84 | 0.75 |
| **5** | 0.71 | 0.70 | 0.81 | 0.71 |
| **6** | 0.70 | 0.65 | 0.86 | 0.75 |
| **7** | 0.68 | 0.70 | 0.81 | 0.73 |
| **8** | 0.63 | 0.76 | 0.83 | 0.69 |
| **9** | 0.67 | 0.77 | 0.81 | 0.75 |
| **10** | 0.59 | 0.71 | 0.77 | 0.78 |
| **Average (± Std.)** | **0.70 (± 0.04)** | **0.70 (± 0.05)** | **0.81 (± 0.03)** | **0.75 (± 0.04)** |

superiority of MSPViT (X2) with a silhouette score of 0.66 (± 0.16) over MSPViT (X1), MSPViT (X3), and MSPViT (X, X1, X2, X3) with AUCs of 0.59 (± 0.18), 0.65 (± 0.20), and 0.65 (± 0.18) respectively, as also concluded in Table 2.13. As mentioned in Table 2.16, MSPViT (X3) with a silhouette score of 0.81 (± 0.03) outperforms MSPViT (X1), MSPViT (X2), and MSPViT (X, X1, X2, X3) with 0.70 (± 0.04), 0.70 (± 0.05), and 0.75 (± 0.04), respectively, in classifying the severity levels. These results are consistent with our findings in Table 2.14 where an F1 score of 96.29% (± 3.01) in severity classification is reported for MSPViT (X3). Comparing Tables 2.15 and 2.16 implies higher dissimilarities among clusters in severity classification than PAD detection. This finding is aligned with the example shown in Fig. 2.24, where mild and severe PAD classes fall apart in the two-dimensional space. The distance between mild and severe PAD classes in the 32-dimensional feature space, while being considered in the same cluster, leads to a lower silhouette score. Additionally, this observation confirms the difference between the nature of mild and severe cases in a high-dimensional space.

### 2.3.4.2 Comparative Analysis

Table 2.17 compares MSPViT with state-of-the-art methods for PAD detection. These methods leverage either blood pressure patterns acquired from an arm cuff [20,21], or PPG sensors placed on toes [23,24,26] to recognize PAD patients. As discussed earlier, state-of-the-art methods employ deep learning (DL) techniques and statistical analysis to process PPG and oscillometry recordings for the detection of PAD.

Our proposed framework for PAD detection, ACM + MSPViT (X2), outperforms other methodologies in terms of sensitivity (99.4% vs. 90.0%, 90.0%, 93.0%, 92.0%, and 82.4% respectively offered by [20], [21], [23], [24], and [26]), implying the lowest missed PAD rate. The specificity of the proposed method, i.e., 98.2%, exceeds the oscillometry method in [20] by a margin of 0.8%, suggesting a lower false detection rate. MSPViT (X2) offers an accuracy of 99.1% which is considerably greater than 94.8%, 91.4%, 91.0%, and 86.9% presented by [20], [21], [23], and [26], respectively. Comparing the F1 score of the

Table 2.17 Comparative Analysis of PAD Detection and Severity Classification Based on Accelerometer Contact Microphone Data Trained on MSPViT Architectures Versus State-of-the-Art

| Methodology | Data Type | Task | Number of PAD Patients | Performance | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Sen (%) | Spec (%) | Acc (%) | F1 score (%) | AUC |
| Oscillometry + DL [20] | Arterial pressure wave | PAD detection | 14 | 90.0 | 97.4 | 94.8 | 92.3 | 0.94 |
| Oscillometry + DL [21] | Arterial pressure wave | PAD detection | 14 | 90.0 | 92.1 | 91.4 | - | - |
| PPG + Statistical Analysis [23] | Bilateral toe PPG | PAD detection | 48 | 93.0 | 89.0 | 91.0 | - | - |
| PPG + Statistical Analysis [24] | Toe PPG | PAD detection | 30 | 92.0 | 45.0 | - | - | 0.79 |
| PPG + DL [26] | Toe PPG | PAD detection | 80 | 82.4 | 89.0 | 86.9 | - | - |
| **ACM + MSPViT (X2)** | **Proximal ACM signals** | **PAD detection** | **74** | **99.4** | **98.2** | **99.1** | **99.3** | **0.99** |
| **ACM + MSPViT (X3)** | **Proximal ACM signals** | **PAD severity classification** | **74** | **96.7** | **97.3** | **97.6** | **96.3** | **0.96** |

proposed method with that reported in [20] (99.3% vs. 92.3%), an improvement of 7.0% is concluded as a result of using ACM proximal recordings with MSPViT (X2). An area under the curve (AUC) of 0.99 by ACM + MSPViT (X2) outperforms the oscillometry method in [20] and the PPG-based method in [24] with AUC scores of 0.94 and 0.79 respectively. Additionally, our proposed methodology is capable of determining the severity level of PAD with an F1 score of 96.3% and average AUC score of 0.96, whereas conventional methods are unable to evaluate the severity level.

## 2.4 MEMS Fingertip Strain Plethysmography for Cuffless Estimation of Blood Pressure

Hypertension, defined as having systolic or diastolic blood pressures (BP) greater than 130 and 80 mmHg respectively [162], is a chronic medical condition affecting approximately 50% of the adult population in the US [163]. Hypertension is the main risk factor for serious health complications such as heart disease [164], stroke [165], and kidney disease [166]. This provokes the need for the early diagnosis of this condition to minimize adverse outcomes and treatment costs.

Strategies proposed for the management of high blood pressure involve low cost per quality-adjusted life year (QALY) [167], demonstrating that the benefits gained from implementing these strategies far outweigh the costs. One suitable strategy would be to have individuals at risk of hypertension check their blood pressure regularly at home. The most common technique for this purpose is using oscillometric cuffs. This involves placing a cuff around the upper arm and gradually deflating the bladder while the device measures blood pressure oscillations and calculates the systolic blood pressure (SBP), the diastolic

blood pressure (DBP), and the mean arterial pressure (MAP). While this method is the only reliable technique for at-home blood pressure monitoring, it can be inconvenient, resulting in a mere 40% of people with hypertension using it at home [168]. Furthermore, accurate readings require strict adherence to guidelines, which only 3% of normal users can successfully achieve [169]. As such, there is a pressing need for convenient and accurate BP monitors for home-based patient monitoring.

A variety of techniques have been proposed over the years for the convenient estimation of blood pressure such as pulse transit time (PTT)-based methods which require two sensors to capture proximal and distal cardiac cycles [61]. This method leverages the inverse relation between BP and PTT alongside a regression model to estimate BP values. Despite promising performance, the regression parameters are subject-specific, hindering the generalization of the model among subjects [61]. On the other hand, the photo-plethysmography (PPG)-driven approach which represents another convenient BP monitoring method, has proven subject invariability when large datasets are used for training the predictive models [170,171]. The tremendous

inclination towards PPG-based methods can be attributed to the common morphological characteristics shared between PPG recordings captured at peripheral arteries and the corresponding continuous arterial blood pressure (ABP) waveforms [172]. As such, huge efforts have been dedicated to the collection of large PPG datasets such as the University of Queensland Vital Signs (UQVS) Dataset [173], the Medical Information Mart for Intensive Care (MIMIC) [174–176], and PPGBP [177], to promote research in PPG-based methods.

There are three primary strategies for estimating blood pressure (BP) from PPG signals. The first set of techniques involves extracting handcrafted features from photoplethysmography (PPG) recordings, with the purpose of estimating blood pressure for a short duration [178–181]. The second set of methods uses deep learning models to link continuous PPG recordings with a single-point blood pressure reading [182,183]. The third group utilizes deep learning to convert PPG signals into their ABP equivalents [184], resulting in a continuous output. Despite great consistency with the standards defined for BP estimation by the US Association for the Advancement of Medical Instrumentation (AAMI), there are two downsides to using PPG sensors, limiting their use in real-world scenarios. Firstly, the optical nature of PPG sensors makes them vulnerable to darker skin tones which tend to absorb light [185]. This inhibits a wide range of the population from appreciating PPG-based BP monitoring technologies and exacerbates health disparities. Second, PPG signals are extremely susceptible to motion artifacts, making them unreliable during routine daily activities [186]. As such, augmenting PPG sensors with modalities that are not sensitive to motion artifacts and darker skin tones is inevitable.

Recent advancements in the development of strain sensors have expanded possibilities for monitoring cardiovascular activities. In [187], a strain sensor was mounted on the radial artery, while a PPG sensor was placed on the fingertip, allowing the determination of blood pressure using pulse transit time. The reported mean absolute errors for the estimated DBP and SBP were 3.71 mmHg and 5.44 mmHg, respectively. In [188], a BP estimation method based on the variation of wrist skin strain in response to changes in blood pressure was proposed. The method was evaluated on 30 subjects, where mean errors

of 1.6 mmHg and -2.0 mmHg were achieved for DBP and SBP, respectively. These studies support the feasibility of using strain sensors for blood pressure monitoring, motivating the present study.

In this work, we take advantage of arterial dilations at the fingertip and establish a novel framework for the estimation of blood pressure. The major contributions of this study are as follows:

- The idea of fingertip strain plethysmography has been introduced for the estimation of blood pressure. The proposed sensing system takes advantage of the fingertip dilation in each cardiac cycle to estimate systolic and diastolic blood pressure values. Furthermore, to the best of our knowledge, this is the first study that addresses the estimation of blood pressure from fingertip pulsation recorded by a strain sensor.
- A novel feature extraction technique based on the morphological characteristics of strain recordings and its first and second-order derivatives and integrals is introduced. The proposed approach initially locates the fiducial points in the signals, and then extracts timing, amplitude, phase, and statistical features of the signal. In addition, a feature set, coined geometric features, is introduced for the first time where the features are extracted based on the trigonometric characteristics of the triangles formed by the fiducial points. These features encode the temporal evolution of the pulse signal, thus enabling the estimation of a wide range of blood pressure values. An ensemble estimator is then utilized to relate the extracted

feature space to the corresponding blood pressure values registered by a standard beat-to-beat monitor.

- Comprehensive analyses are conducted to quantify the contribution of each feature to the estimation of BP. The robustness of the proposed framework is also evaluated against body mass index (BMI) and blood pressure level. In addition, an ablation study is performed to determine the impact of the absence of each feature category on the estimation errors for DBP, SBP, and MAP.

### 2.4.1 Arterial Dilation in Fingertip

Finger distension, also known as finger pulsation or finger dilation, is a common phenomenon resulting from the heartbeat. When the heart contracts, it propels blood into the arterial system, causing a pressure wave to spread throughout the blood vessels. This pressure wave is palpable in various parts of the body, including the index finger as depicted in Fig. 2.25. As can be seen, the digital artery is responsible for directing oxygen-rich blood to the finger tissue. Figs 2.25 (a) and (b) illustrate the shape of the index finger before and after blood flow through the digital artery, respectively. It has been shown that finger dilation can stretch or compress a strain sensor, thereby altering its capacitance [189]. The time evolution of these changes, therefore, indicates the heartbeat-induced pressure wave in the artery.

As discussed earlier, PPG signals have gained ground for estimating blood pressure due to their similarity with arterial blood pressure signals at peripheral sites. In Fig. 2.25 (c), the morphology of the strain recording from the fingertip is compared with a simultaneously-recorded finger PPG signal obtained using an optical sensor. The

consistent peaks and valleys observed in both signals suggest that strain sensors hold great potential for blood pressure monitoring, without being affected by optical limitations. This study explores the capability of strain sensors for blood pressure monitoring.

### 2.4.2 Experimental Setup and Methodology

#### 2.4.2.1 Data Collection and Study Protocol

The study was conducted on 32 subjects, 26 males and 6 females, with average (± standard deviation) ages of 25 (± 3.28) and 23.16 (± 0.75), respectively. Among all 32 subjects, 26 had normal blood pressures while 6 were identified as either with elevated

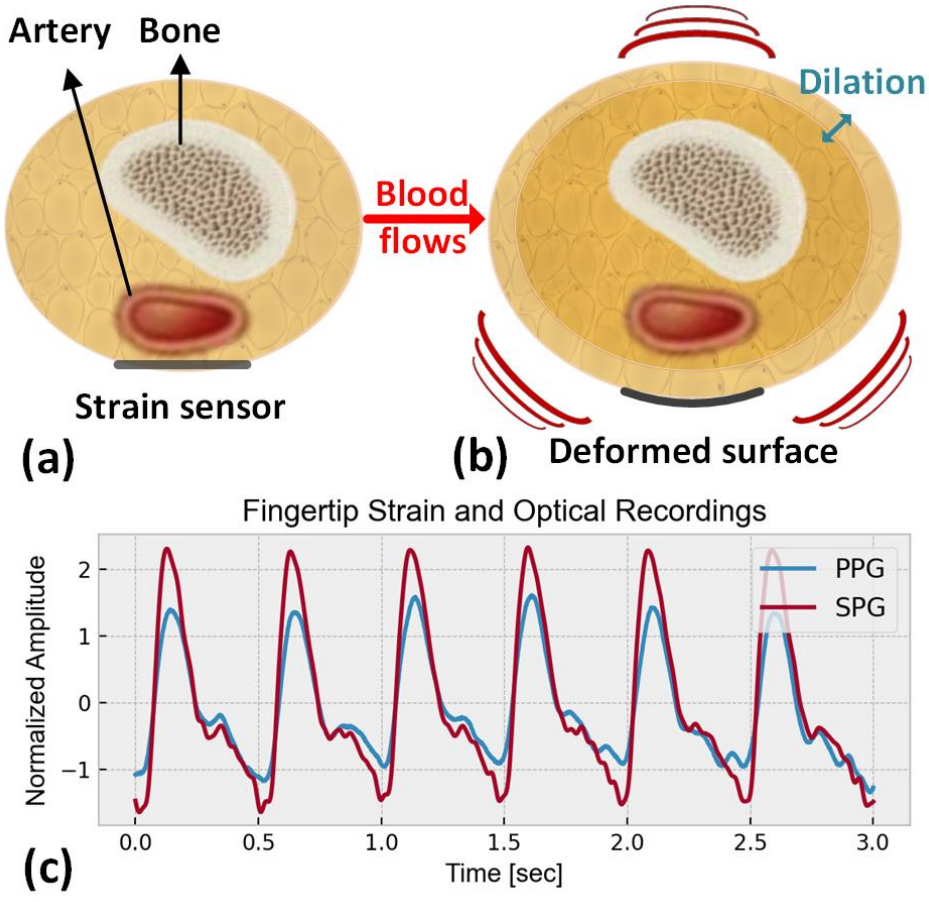


Fig. 2.25 Strain plethysmography of fingertip. (a) a normal finger, (b) a dilated finger due to the blood flow, and (c) comparison between strain and optical recordings of the fingertip.

Table 2.18 Demographic Information of the Participants

| Gender | Age (years) | Weight (kg) | Height (cm) | BMI (kg/m2) | SBP (mmHg) | DBP (mmHg) | Heart Rate (BPM) |
|---|---|---|---|---|---|---|---|
| **Male (26)** | 25.0 ± 3.28 | 75.35 ± 9.08 | 175.88 ± 5.73 | 25.02 ± 3.26 | 125.15 ± 14.69 | 81.92 ± 9.87 | 84.07 ± 12.42 |
| **Female (6)** | 23.16 ± 0.75 | 64.61 ± 16.43 | 162.16 ± 11.44 | 23.55 ± 4.05 | 107.83 ± 8.15 | 74.33 ± 3.26 | 87.50 ± 11.37 |
| **Total (32)** | 24.65 ± 3.05 | 74.65 ± 12.98 | 173.31 ± 8.78 | 24.74 ± 3.40 | 121.90 ± 15.23 | 80.50 ± 9.45 | 84.71 ± 12.13 |

blood pressure (130 mmHg < SBP < 140 mmHg) or having hypertension (SBP > 140 mmHg). The average (± standard deviation) systolic and diastolic blood pressures among all subjects were 121.90 (± 15.23) mmHg and 80.50 (± 9.45) mmHg, respectively. Further demographic information on the subjects is summarized in Table 2.18.

Fig. 2.26 demonstrates the suggested framework for blood pressure estimation. The experimental setup, as shown in Fig. 2.26 (a), includes a custom-built, sensitive capacitive micro-electromechanical systems (MEMS) strain sensor that is attached to the tip of the subject's index finger using medical-grade adhesive tape. The attachment is depicted in Fig. 2.26 (b). We have used a wafer-level-packaged capacitive MEMS strain sensor for fingertip SPG. The sensor utilizes a parallel plate capacitor with a 200nm gap size to achieve high sensitivity. The sensor design is shown in Fig. 2.26 (c). The strain sensitivity of the sensor is 73 fF/με, corresponding to $8.5 \times 10^3$ ppm/με (Fig. 2.26 (d)). Compared to other MEMS strain sensors reported in literature [190], this device shows more than 100 times larger sensitivity and a much greater signal-to-noise ratio. This sensor was utilized to register minute fingertip dilations at a rate of 1 kHz. Additionally, a tonometry-based TL-300 from Tensys Medical was mounted to the subject's wrist to record reference blood pressure values from the radial artery on a second-by-second basis. The experimental setup

also involves either a three-lead ECG sensor node (ECG Development Kit; Shimmer Sensing, Dublin, Ireland) with the electrodes attached to the right arm (RA), left arm (LA), right leg (RL), and left leg (LL), or a photoplethysmography (PPG) sensor (Shimmer3 GSR; Shimmer Sensing, Dublin, Ireland) attached to the earlobe using an ear clip.

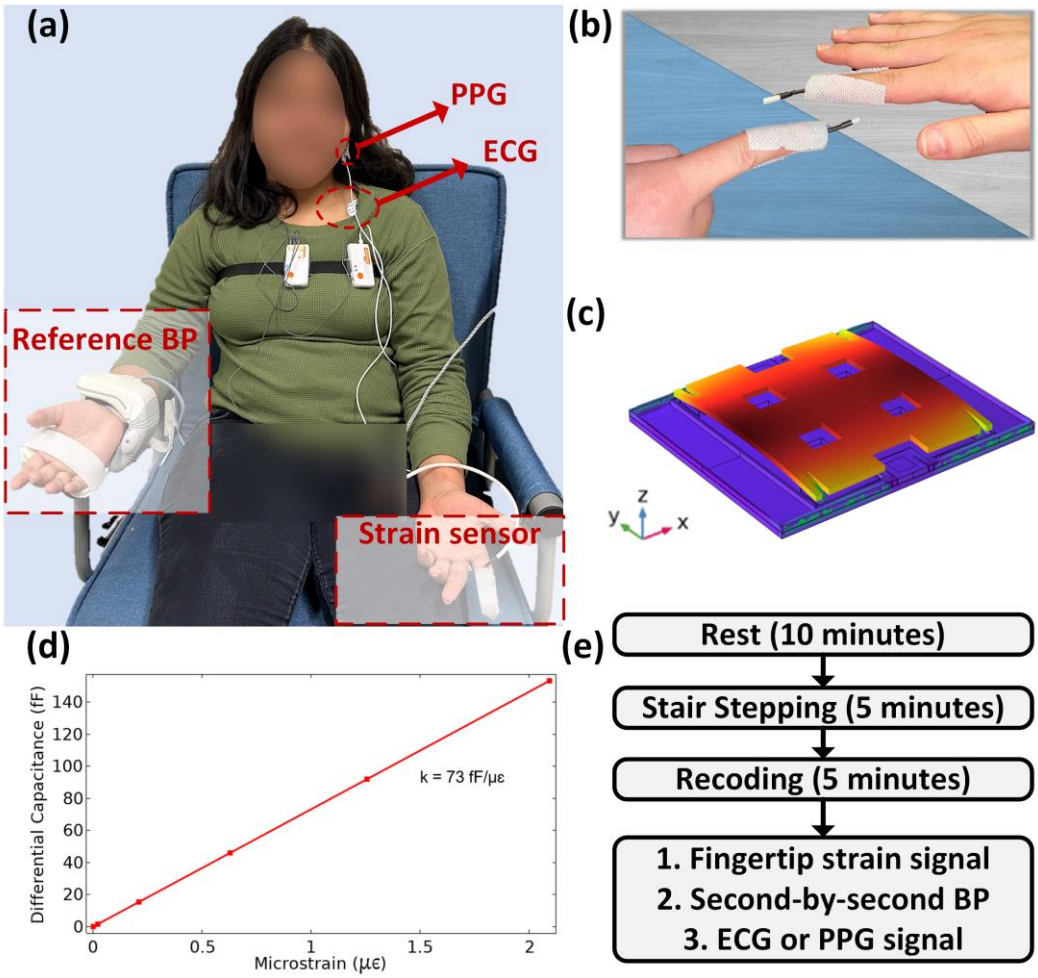


Fig. 2.26 The proposed blood pressure estimation framework. (a) The arrangement of the fingertip strain sensor, ECG, PPG, and reference BP on the subject, (b) attachment of the strain sensor to the fingertip from two different angles, (c) the mode shape of the MEMS strain sensor, (d) the scale factor/sensitivity of the sensor is 73 fF/με, and (e) measurement protocol.

Both ECG and PPG sensors are shown in the figure for demonstration purposes. The measurement protocol is demonstrated in Fig. 2.26 (e). The ECG and PPG data were streamed with a sampling frequency of 512 Hz. Initially, each subject was seated at rest on a chair for at least 10 minutes, after which their baseline blood pressure was measured using an OMRON-BP786N (a standard cuff-based blood pressure monitor). The subject was then asked to perform 5 minutes of stair stepping which temporarily increases the blood pressure. Immediately after, the subject was seated on a chair for 5 minutes, during which the fingertip strain signal, reference blood pressure, and either ECG or PPG signal were recorded using the aforementioned setup. The subject experimental protocol was approved by the Institutional Review Board of Stevens Institute of Technology under protocol number 2022-044 (N), based on which informed consent was obtained from the subjects. It should be noted that this study does not utilize ECG and PPG data and these signals were recorded for future evaluations. In the proposed framework, the objective is to estimate blood pressure using short segments of fingertip strain recordings.

#### 2.4.2.2 Data Preparation

To remove baseline wander and noise components, a zero-phase $2^{nd}$ order Butterworth filter was used to bandpass-filter the strain recordings within the range of 0.8-25 Hz. This frequency is commonly used for filtering PPG signals [191], which share a common nature with fingertip strain signals. The filtered signals were then synchronized with the reference BP time series using the respective timestamps. The recording of each subject was then segmented into 3-second non-overlapping segments. In parallel, blood pressure time series including systolic, diastolic, and mean arterial pressure were averaged

over 3-second intervals to generate labels corresponding to strain segments. Segments with extremely distorted reference values, i.e., pressure out of 30-220 mmHg or pulse pressure (difference between systolic and diastolic) smaller than 10 mmHg, were discarded, resulting in a reliable dataset.

#### 2.4.2.3 Fingertip Strain Signal Annotation

The characterization of the fingertip strain plethysmography signals begins with annotation of the waveform with feature points shown in Fig. 2.27. For this purpose, the 1st-order derivative (gradient; $d(t)$) and 2nd-order derivative (curvature; $dd(t)$) of each fingertip strain segment are calculated to highlight the dynamics of the waveform. Additionally, the 1st and 2nd-order integrals ($i(t)$ and $ii(t)$ respectively) are calculated to quantify signal power within the cardiac cycle while discarding noise components.

Fig. 2.27 (c) depicts a 3-second segment of the fingertip strain recording ($S(t)$) which contains three cardiac cycles. As observed, each cardiac cycle contains two local maximum points called systolic ($S_2$) and diastolic ($S_4$) peaks, respectively. In order to detect the systolic peaks, the negative values in each 3-second segment are set to zero, and the resulting signal is squared to increase the prominence of the systolic peaks compared to noise and low-frequency components. The algorithm then computes two moving averages of the squared signal, one with a window length of 110 ms (110 samples) and the other with a window length of 667 ms (667 samples) resulting in $MA_1(t)$ and $MA_2(t)$ respectively. These are used to find the baseline of the strain signal between the beats and

within smaller durations, respectively. A threshold is then defined as follows, with which $MA_1(t)$ is then compared:

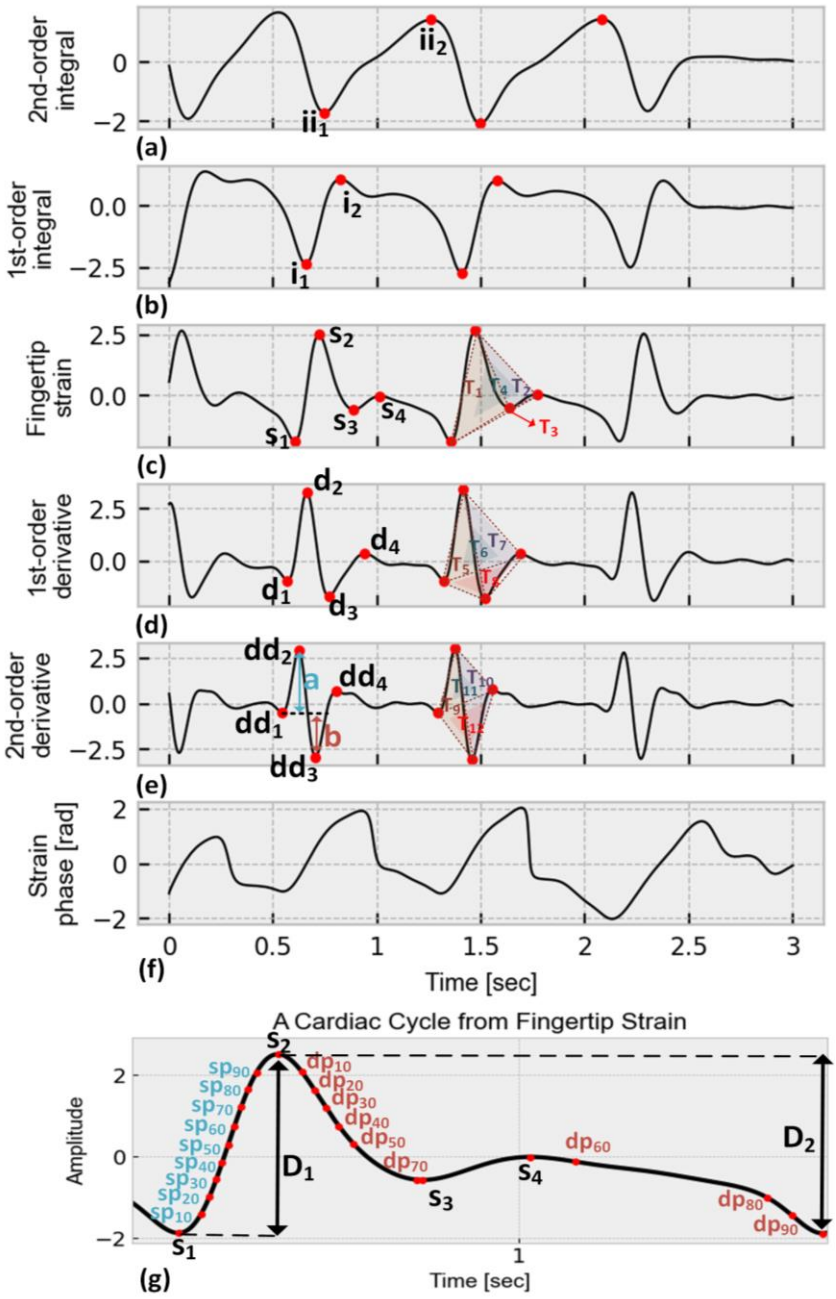


Fig. 2.27 The visual presentation of fingertip strain recordings. (a) 2nd-order integral, (b) 1st-order integral, (c) time evolution, (d) 1st-order derivative, (e) 2nd-order derivative, (f) phase, and (g) cardiac cycle of the fingertip strain signal.

$$\gamma(t) = MA_2(t) + 0.02 \times \overline{S^2(t)}, \qquad (36)$$

where $\overline{S^2(t)}$ denotes the average value of the squared segment. The points in $MA_1(t)$ exceeding the threshold are identified as systolic peaks in $MA_1(t)$. Finally, the algorithm identifies the systolic peaks by finding the local maxima in the original strain signal and enforcing a minimum delay of 480 ms (480 samples) between the detected peaks to discard peaks which are too close to each other.

The local minima within 200 ms preceding the systolic beat in a cardiac cycle are marked as the valley points ($S_1$). Fig. 2.27 (d) demonstrates the gradient of the strain segment which is represented by two local maxima ($d_2$ and $d_4$) and one minimum ($d_3$) in each cardiac cycle. The local maximum between $S_1$ and $S_2$ is annotated with $d_2$ which is followed by a minimum ($d_3$) between $S_2$ and the following $S_2$. The local maximum within the range of $d_3$ and the following $S_1-5$ ms signifies $d_4$. As illustrated in Fig. 2.27 (e), the curvature of the fingertip strain signal is comprised of four feature points, $dd_1$, $dd_2$, $dd_3$, and $dd_4$, representing the minima before $dd_2$, local maxima within $S_1-S_2$, local minima within $dd_2-d_4$, and local maxima within $dd_3-d_4$, respectively. Figs 2.27 (a) and (b) respectively show the 2nd and 1st-order integrals of the fingertip strain signal. The minima ($i_1$) and maxima ($i_2$) associated with the 1st-order integral are located as the lowest and highest values within $S_1 - S_2$ and the range between $i_1$and the following $S_1$, respectively. In each cardiac cycle of the 2nd-order integral of the strain signal, the local minima and maxima are represented by $ii_1$ and $ii_2$, respectively, which correspond to the ranges $i_1 - i_2$

and $i_1 - S_1$ located in the following cycle. The diastolic peak ($S_4$) shown in Fig. 2.27 (c) can then be determined by finding the maxima between $d_3$ and $ii_2$.

Similar to PPG signals, fingertip strain signals exhibit the dicrotic notch which is caused by the closure of the aortic valve during the diastolic phase of the cardiac cycle. This closure causes a rapid increase in the aortic pressure, leading to a small pressure wave which is reflected back to the heart [192]. As shown in Fig. 2.27 (c), the dicrotic notch ($S_3$) can be detected by choosing the point with the lowest slope between $S_2$ and $S_4$ in the strain signal. However, the dicrotic notch can be less visible in certain cases such as in individuals with increased arterial stiffness. In these cases, a straight line connecting $S_2$ and $S_4$ is subtracted from the signal, and the minimum value is selected as the dicrotic notch of the cycle. In addition to the above-mentioned feature points, the points holding an amplitude of X% ($X \in \{10, 20, 30, \dots, 90\}$) of $D_1 = S(S_2) - S(S_1)$ and $D_2 = S(S_1) - S(S_2)$ are determined and denoted by $sp_X$ and $dp_X$, respectively, as depicted in Fig. 2.27 (g). The fiducial points will be leveraged to represent a variety of characteristics in the fingertip strain signal, as described in the following section.

#### 2.4.2.4 Feature Space

##### 2.4.2.4.1 Amplitude features

Amplitude features capture variations in the magnitude of the waveform. The amplitude of the fingertip strain signal is related to the arterial wall force imposed by the blood flow, and changes in this force can be indicative of various physiological processes. Hence, the amplitudes of $S(t)$, $d(t)$, $dd(t)$, $i(t)$, and $ii(t)$ were determined in all 34 feature points discussed in the previous section, yielding 170 features ($5 \times 34$).

#### 2.4.2.4.2 Timing features

It has been demonstrated that temporal characteristics in PPG signals suggest correlation with blood pressure values. Similarly, timing intervals are expected to contribute to the estimation of blood pressure in fingertip strain recordings. As such, the time difference between each pair of the annotated 34 points were calculated, resulting in $C(34, 2) = 561$ timing features.

#### 2.4.2.4.3 Phase-related features

As fingertip strain recordings have only real parts, the signals initially underwent a Hilbert transform to obtain the analytic counterparts. Subsequently, the phase information was achieved using the $tan^{-1}\{.\}$ operator followed by a phase unwrapping within $(-\pi, +\pi)$, as illustrated in Fig. 2.27 (f). The periodic nature of the strain phase suggests that it contains useful information about cardiac events. Hence, the phase values of all 34 annotated points were calculated from the phase signal and added to the feature space.

#### 2.4.2.4.4 Geometric features

As observed in the collected data, blood pressure level may affect morphological characteristics, such as the rise time from $S_1$ to $S_2$ and the sharpness of the systolic and diastolic peaks. As such, we encode the relative locations of the feature points by forming 12 triangles including T1 ($S_1$, $S_2$, $S_3$), T2 ($S_2$, $S_3$, $S_4$), T3 ($S_1$, $S_3$, $S_4$), T4 ($S_1$, $S_2$, $S_4$), T5 ($d_1$, $d_2$, $d_3$), T6 ($d_1$, $d_2$, $d_4$), T7 ($d_2$, $d_3$, $d_4$), T8 ($d_1$, $d_3$, $d_4$), T9 ($dd_1$, $dd_2$, $dd_3$), T10 ($dd_2$, $dd_3$, $dd_4$), T11 ($dd_1$, $dd_2$, $dd_4$), and T12 ($dd_1$, $dd_3$, $dd_4$), as illustrated in Figs 2.27 (c), (d), and (e). The sides and angles of the 12 triangles were determined using the sine and cosine laws, introducing a total of 72 additional features ($6 \times 12$) to the feature space.

#### 2.4.2.4.5 Statistical features

Predictive models are generally trained to learn the distribution of the data. Hence, we take advantage of statistical parameters, such as mean, median, variance, standard deviation, skewness, and kurtosis (6 features), to introduce the amplitude of strain recordings to the predictive model.

#### 2.4.2.4.6 Demographic features

Recent studies have demonstrated that demographic information such as gender, age, height, weight, and BMI are related to a subject's blood pressure value. As such, alongside the prepared feature space, we incorporated demographic information (5 features).

#### 2.4.2.4.7 Miscellaneous features

Five well-known features including the augmentation index (AI; the ratio of the height of the diastolic peak to the height of the systolic peak), large artery stiffness index (LASI; the proportion of the subject's height to the time it takes the pressure wave to travel from the heart to the peripheral sites and back), the ratio of the peak of $dd_3$ to the peak of $dd_2$ (b/a), systolic area (from $S_1$ to $S_2$), and diastolic area (from $S_4$ to the following $S_1$) were added to the feature space. The inclusion of these features has been demonstrated to enhance BP estimation results [193].

It is to be noted that the features calculated for each cardiac cycle, such as time intervals, are averaged across all beats within the same 3-second segment to generate a single value. As a result, each 3-second strain segment is represented by an 853-element vector (170 + 561 + 34 + 72 + 6 + 5 + 5 features) to be correlated with blood pressure.

#### 2.4.2.5 Predictive Model for Blood Pressure Estimation

To estimate blood pressure from fingertip strain segments, we need to find the relationship between the extracted features from 3-second strain segments and blood pressure. This relationship can be defined as the mathematical function:

$$M: F \rightarrow \{\widehat{\mathrm{SBP}}, \widehat{\mathrm{DBP}}, \text{ or } \widehat{\mathrm{MAP}}\}, \qquad (37)$$

where $M$, $F$, $\widehat{\mathrm{SBP}}$, $\widehat{\mathrm{DBP}}$, and $\widehat{\mathrm{MAP}}$ denote the mapping function, input feature space, estimated systolic blood pressure, estimated diastolic blood pressure, and mean arterial pressure, respectively. A regression model can then be adopted to perform the transformation of feature vectors to blood pressure values. The relationship between the morphological features of the pulse and blood pressure is not necessarily linear [193]. Therefore, the regressor model adopted for estimation should be capable of non-linear modeling. To this end, we employ the eXtreme Gradient Boosting (XGBoost) regressor, an ensemble algorithm which combines multiple decision trees to make accurate predictions. The training procedure will be explained in the following section.

### 2.4.3 Experimental Results

This section provides an overview of the training procedure, the evaluation metrics used to measure accuracy, the performance of the model for blood pressure estimation, and the most contributing features for blood pressure estimation.

#### 2.4.3.1 Training Procedure

As mentioned earlier, each fingertip strain segment is represented by 595 features which signify the input to the regression model, while the output represents the estimated blood pressure value. As such, the XGBoost regressor is trained in a supervised manner to

learn and estimate the blood pressure values from the extracted features. The objective of the training procedure is to minimize the following loss function:

$$\ell(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \beta\|w\|_1, \qquad (38)$$

where $y_i$ and $\hat{y}_i$ denote the ground truth and predicted blood pressure values (either SBP, DBP, or MAP), respectively. $\beta\|w\|_1$ represents the lean absolute shrinkage and selection operator (LASSO) regularization term, with $\beta$ and $\|w\|_1$ respectively showing the regularization parameter and the Manhattan norm of the model weights used to prevent overfitting. The training procedure was conducted using the data of 32 subjects, amounting to 3,302 fingertip pulse segments (2 hours and 45 minutes) collected by our MEMS strain

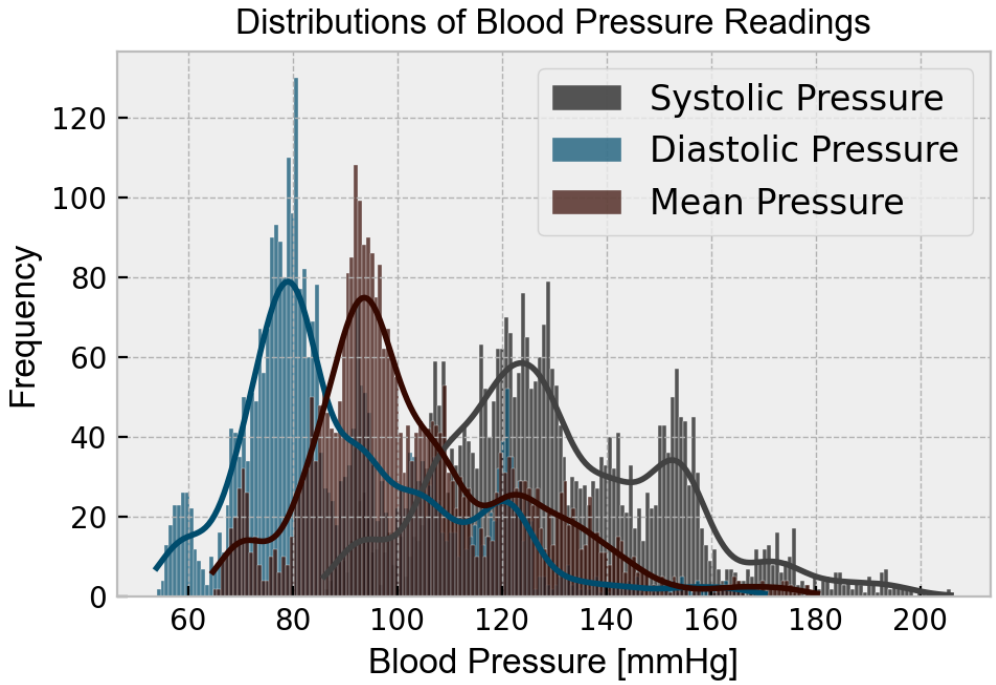


Fig. 2.28 The distribution of systolic, diastolic, and mean blood pressure readings collected by the reference sensor from 32 subjects.

plethysmography sensor. As illustrated in Fig. 2.28, the ground-truth SBP, DBP, and MAP values are distributed within the ranges of [85, 220] mmHg, [54, 160] mmHg, and [65, 180] mmHg, respectively. A 5-fold cross-validation was adopted to assess the generalizability of the model to unseen data. To this end, the dataset was split into train (2,642 segments; 80%) and test (660 segments; 20%) sets. The outliers within each set were identified based on the Z-score, and discarded if the score exceeded 3. The model was implemented on a 24-GB NVIDIA GeForce RTX 3090 Ti FTW3 GPU and trained on the train set. The performance was reported using the test set when the loss function defined in (38) was minimized.

#### 2.4.3.2 Evaluation Metrics

Evaluation of the regression model based on appropriate metrics is essential to measure the accuracy of blood pressure estimations. In this study, we follow the standard metrics defined by AAMI/ESH/ISO [194], including mean absolute error (MAE), mean error (ME), mean error standard deviation (SD), and root mean square error (RMSE) as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \qquad (39)$$

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i), \qquad (40)$$

$$\mathrm{SD} = \sqrt{\frac{1}{N-1}\times\sum_{i=1}^{N}(y_i - \hat{y}_i - \mathrm{ME})^2}, \qquad (41)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \qquad (42)$$

which indicate the accuracy of estimation regardless of whether the prediction is too high or low, the amount of bias in the estimated BP values and the corresponding standard deviations, and the estimation accuracy with emphasis on large errors of BP, respectively. Hence, smaller errors are suggestive of more accurate estimations.

Furthermore, the estimation accuracy is statistically assessed based on the coefficient of determination, which is also known as R-squared ($R^2$) value. This metric indicates the proportion of variance in the estimated blood pressure values that is predictable by the corresponding ground-truth BP values, as defined by:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \qquad (43)$$

where $\bar{y}$ denotes the average of ground-truth BP values. The numerator in (43) shows the error of the estimated BP values whereas the denominator represents the deviation from the mean within the actual BP set. As such, a perfect estimation would result in an $R^2$ of 1 (i.e., an error of zero) while $R^2 = 0$ suggests the most inaccurate estimation.

#### 2.4.3.3 BP Estimation Performance

The performance of the methodology is summarized for two training approaches in Table 2.19. The first row corresponds to the 5-CV approach described earlier, while the second row presents the results of subject-specific training procedure in which the data of the same subject constitutes both the training and the test sets (80:20). The third row summarizes the average of the two training approaches. As observed for 5-CV, RMSEs of 8.03, 6.12, and 6.50 mmHg were achieved for SBP, DBP, and MAP, respectively, denoting

Table 2.19 Performance Evaluation for the Estimation of SBP, DBP, and MAP

| Training approach | SBP | | | | DBP | | | | MAP | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | $R^2$ | MAE | ME ± SD | RMSE | $R^2$ | MAE | ME ± SD | RMSE | $R^2$ | MAE | ME ± SD |
| Pooled data (5-CV) | 8.03 | 0.86 | 5.49 | -0.02 ± 8.03 | 6.12 | 0.90 | 3.88 | -0.08 ± 6.12 | 6.50 | 0.89 | 4.25 | -0.14 ± 6.50 |
| Subject-specific | 8.02 | 0.85 | 5.03 | 0.01 ± 8.02 | 6.08 | 0.90 | 3.70 | 0.24 ± 6.08 | 5.83 | 0.91 | 3.83 | 0.18 ± 5.83 |
| **Average (± std.)** | **8.03** | **0.86** | **5.26** | **-0.01 ± 8.01** | **6.10** | **0.90** | **3.84** | **0.08 ± 6.10** | **6.16** | **0.90** | **4.04** | **0.02 ± 6.17** |

a small average deviation of 6.88 (± 0.82) mmHg from ground-truth values. The highest predictable variance was achieved in the estimated DBP values with an $R^2 = 0.90$, whereas the lowest $R^2$ of 0.86 belongs to the estimated SBPs. The small variation from MAP to SBP demonstrates the robustness of the estimator to a variety of BP ranges. The MAEs reported for SBP, DBP, and MAP (5.49, 3.88, and 4.25 mmHg respectively) satisfy the 5-mmHg difference recommended by AAMI to a large extent. The MEs achieved for SBP, DBP, and MAP signify average biases of -0.02, -0.08, and 0.14 mmHg, respectively, implying excellent predictive capabilities. Subject-specific analyses have resulted in similar performance to 5-CV with RMSE, $R^2$, MAE, and ME ranges of [5.83, 8.02] mmHg, [0.85, 0.91], [3.7, 5.03] mmHg, and [0.01, 0.18] mmHg, respectively, demonstrating the reliability of the estimator for a wide range of blood pressure values regardless of the training approach.

Statistical analyses were conducted to further assess the BP estimation performance. Figs 2.29 (a), (b), and (c) show the scatter plots of the estimated SBP, DBP, and MAP in terms of the ground-truth values, respectively, as well as the identity function ($y = x$). The points are distributed in the 2D planes of SBP, DBP, and MAP with slopes of 0.997, 1.016, and 1.011, respectively, indicating linear relationships between actual and

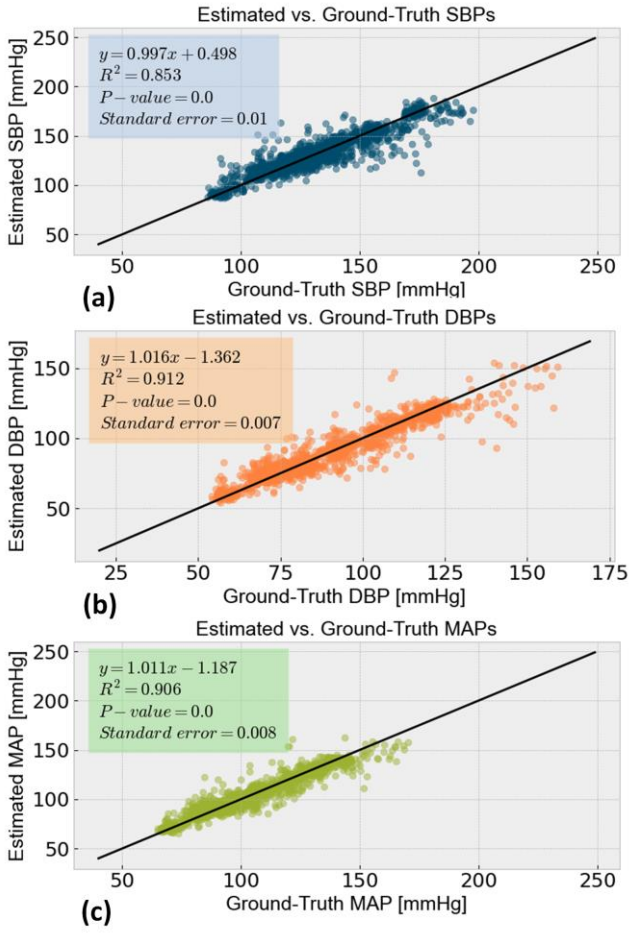


Fig. 2.29 Scatter plots of the estimated (a) SBP, (b) DBP, and (c) MAP values in 5-CV analysis, in terms of the respective ground-truth values.

estimated values. The small standard errors of 0.01, 0.007, and 0.008 mmHg for SBP, DBP, and MAP, respectively, under the assumption of residual normality demonstrate low uncertainty and error in the estimated values. The t-test shows extremely small p-values (approximately 0) for all three blood pressure types, deducing significant correlations between the estimated and ground truth BP values.

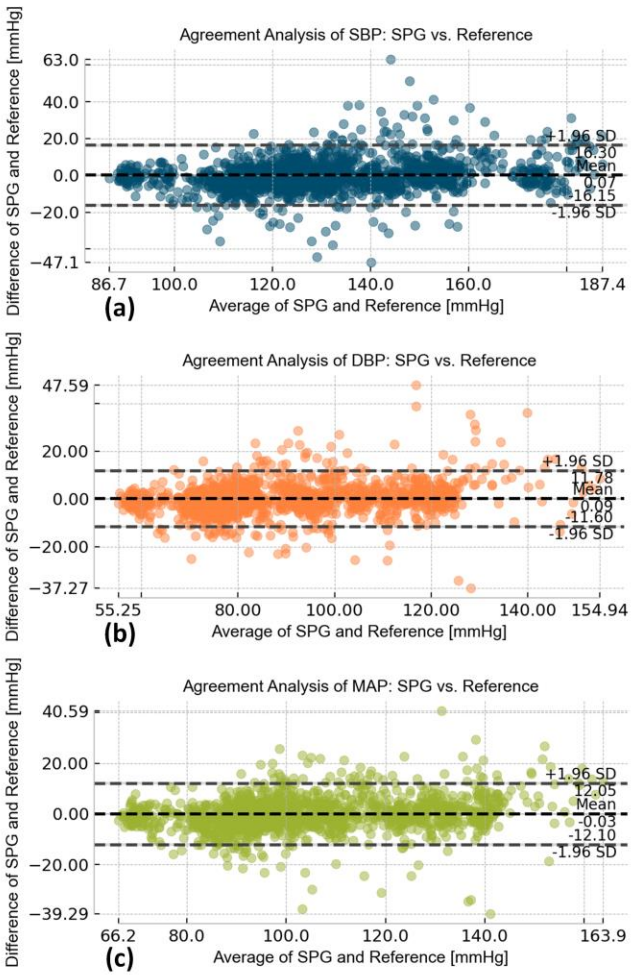


Fig. 2.30 The Bland-Altman agreements between the estimated and reference (a) SBP, (b) DBP, and (c) MAP values in 5-CV analysis.

The statistical agreement between the estimated and reference SBP, DBP, and MAP values are investigated in Bland-Altman characteristics with confidence intervals of 95% (±1.96 × SD) as depicted in Figs 2.30 (a), (b), and (c), respectively. According to these figures, the majority of differences for SBP, DBP, and MAP fall within the agreement limit ranges of [-16.15,16.30], [-11.60,11.78], and [-12.10,12.05] mmHg, respectively.

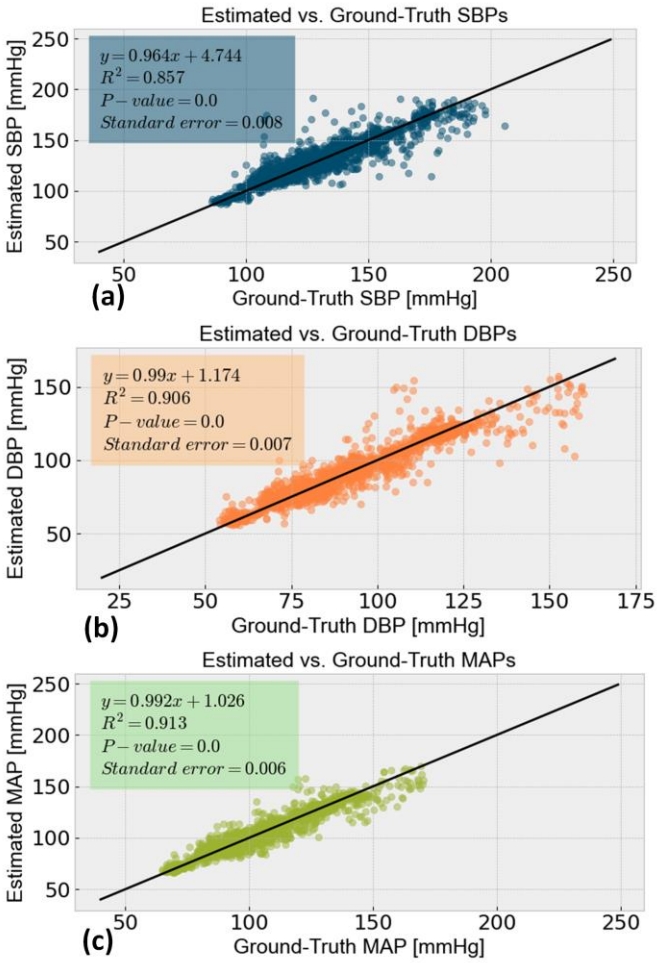


Fig. 2.31 Scatter plots of the estimated (a) SBP, (b) DBP, and (c) MAP values in subject-specific analysis, in terms of the respective ground-truth values.

Furthermore, the mean differences of 0.07, 0.09, and -0.03 mmHg for SBP, DBP, and MAP, respectively, suggest negligible estimation biases.

In a similar manner, the scatter plots and Bland-Altman characteristics of the subject-specific model are summarized in Figs 2.31 and 2.32, respectively. Small standard errors of 0.008, 0.007, and 0.006 for SBP, DBP, and MAP are observed, respectively. These along with high $R^2$ values (0.857, 0.906, and 0.913 for SBP, DBP, and MAP

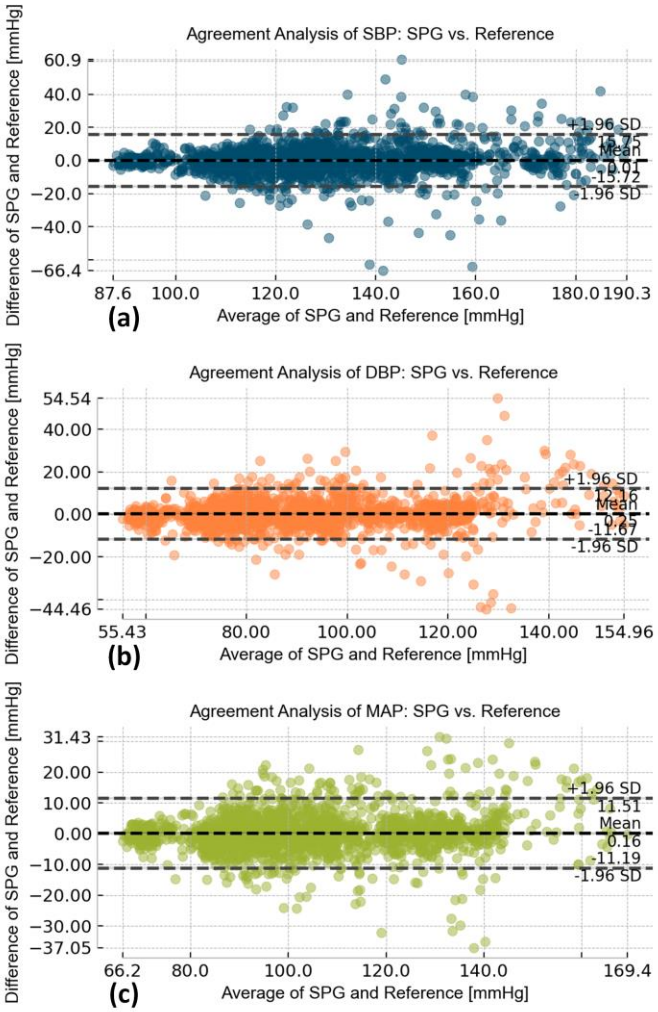


Fig. 2.32 The Bland-Altman agreements between the estimated and reference (a) SBP, (b) DBP, and (c) MAP values in the subject-specific analysis.

respectively) imply superior estimation accuracy. As shown in Fig. 2.32 (a), the difference between the estimated and reference SBP values are majorly within the range of [-15.72,15.75] BPM with an average bias of 0.07 mmHg, while the differences are within smaller agreement limit ranges for DBP ([-11.60,11.72] mmHg) and MAP ([-12.10,12.05] mmHg) as illustrated in Figs 2.32 (b) and (c), respectively.

### 2.4.3.4 Contributing Features to BP Estimation

As demonstrated in the previous section, the 853-dimensional feature space could fairly represent wide ranges of systolic, diastolic, and mean pressure. This section aims to identify the most contributing features to the estimation of blood pressure. For this purpose, we utilized the gains attributed to each feature during the construction of the decision trees in the ensemble model. The gains achieved from the five folds are then averaged, to represent the contribution of the feature. Figs 2.33 (a), (b), and (c) rank the top 30

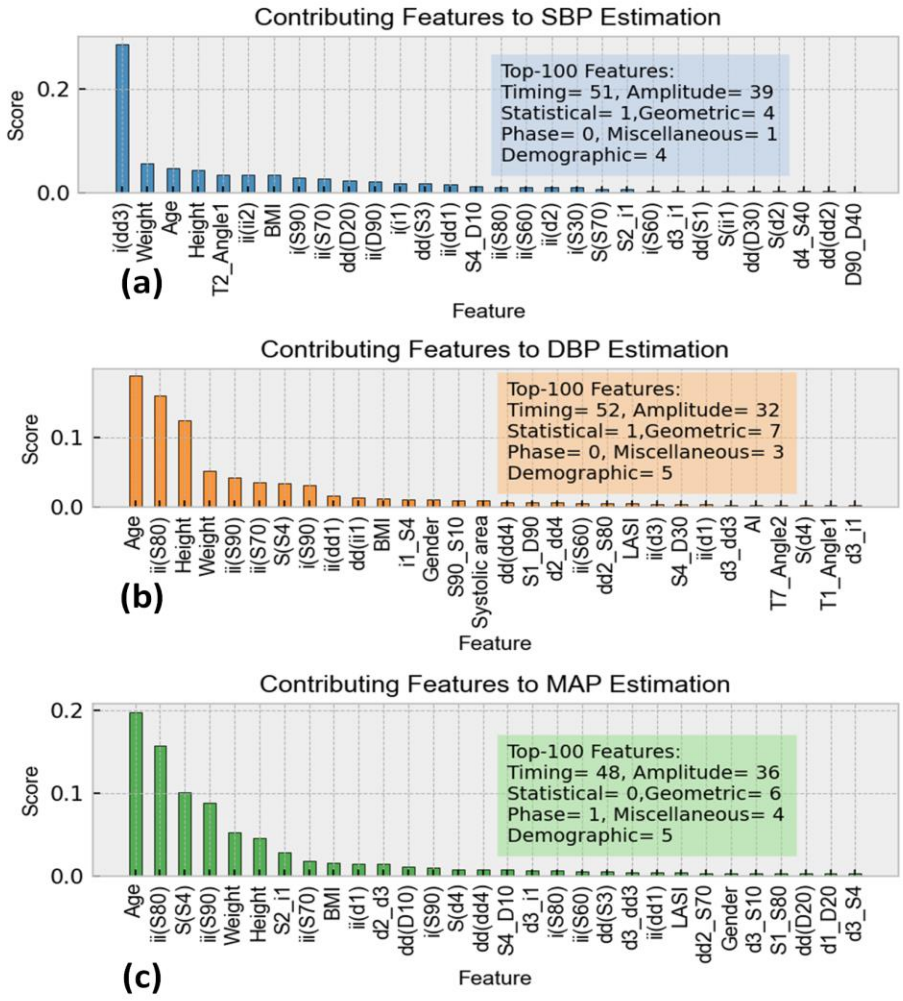


Fig. 2.33 The top 30 contributing features to the estimation of (a) SBP, (b) DBP, and (c) MAP.

contributing features to the estimation of SBP, DBP, and MAP, respectively. Among the top-30 lists, 15, 11, 10 features belong to the integral signals, which are introduced in the present study. Additionally, "T2_Angle1" which shows the angle associated with $S_2$ (systolic peak) is ranked $4^{th}$ among the top features for estimating SBP. This finding verifies the observed difference in the sharpness of the systolic peak due to blood pressure variations. The "T1_Angle1" and "T7_Angle2" which mark the angles of $S_1$ and $d_3$ in the T1 and T7 triangles of the strain segment and first-derivative signal, respectively, are ranked among the top-30, demonstrating the significance of the geometric features. The presence of 4, 5, and 5 demographic features in the top-30 lists associated with SBP, DBP, and MAP demonstrates that the combination of demographic information along with the introduced geometric and integral-based features could precisely represent blood pressure values. Meanwhile, the contribution of each feature category to the top 100 features is listed in a shaded box within respective figures. As can be seen, timing features with the counts of 51, 52, and 48 are the most contributing features to the estimation of SBP, DBP, and MAP, respectively, whereas amplitude and geometric features are ranked in the second and third places. Despite the extraction of phase information of 34 feature points, they have no contribution to blood pressure estimation (except for only one feature which contributes to MAP).

### 2.4.4 Discussion

In the previous section, we comprehensively investigated the feasibility of estimating BP from fingertip strain plethysmography recordings. This section examines the impact of blood pressure level and body mass index (BMI) of the subjects on the BP

estimation accuracy. Furthermore, an ablation study on the feature space, a comparative analysis with state-of-the-art, and the limitations of the present study as well as future directions will be discussed.

#### 2.4.4.1 Impact of Blood Pressure Level

Human blood pressure (SBP/DBP) can range from below 90/60 mmHg (hypotension) to above 130/80 mmHg (hypertension). It is therefore, essential for a BP monitor to be able to provide accurate estimations across wide ranges. The robustness of SPG-based BP monitoring was evaluated against BP levels of normal (SBP/DBP < 120/80 mmHg), elevated (120/80 mmHg < SBP/DBP < 130/80 mmHg), stage I hypertension (130/80 mmHg < SBP/DBP < 140/90 mmHg), stage II hypertension (140/90 mmHg < SBP/DBP < 180/120 mmHg), and crucial (SBP/DBP > 180/120 mmHg), as summarized in Table 2.20. The RMSE values show ranges of [2.35, 6.35] mmHg, [1.82, 6.69] mmHg, and [1.75, 5.71] mmHg for SBP, DBP, and MAP across all subjects, respectively, indicating excellent performance for a wide range of BP values. The lowest RMSE belongs to the elevated group with 2.35, 1.82, and 1.75 mmHg for SBP, DBP, and MAP

Table 2.20 The Impact of Blood Pressure Category on Estimation Performance

| **BP Category (segments)** | **SBP** | | | **DBP** | | | **MAP** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **RMSE** | **ME ± SD** | **MAE** | **RMSE** | **ME ± SD** | **MAE** | **RMSE** | **ME ± SD** | **MAE** |
| Normal (852; 25%) | 4.11 | 0.01 ± 4.11 | 3.07 | 3.23 | 0.02 ± 3.23 | 2.39 | 3.14 | 0.05 ± 3.15 | 2.39 |
| Elevated (295; 9%) | 2.35 | 0.06 ± 2.35 | 1.88 | 1.82 | 0.16 ± 1.82 | 1.28 | 1.75 | 0.13 ± 1.75 | 1.35 |
| Stage I (956; 29%) | 6.12 | -0.06 ± 6.12 | 3.99 | 3.77 | -0.06 ± 3.77 | 2.63 | 3.94 | 0.05 ± 3.94 | 2.77 |
| Stage II (1,192; 35%) | 6.35 | 0.09 ± 6.35 | 4.28 | 6.69 | 0.05 ± 6.69 | 4.23 | 5.71 | 0.04 ± 5.71 | 3.86 |
| Critical (58; 2%) | 5.18 | 0.02 ± 5.23 | 4.08 | 5.84 | 1.17 ± 5.77 | 3.58 | 5.28 | 0.91 ± 5.25 | 3.61 |
| **Average (± std.)** | **4.82 (± 1.46)** | **0.02 (± 4.76)** | **3.45 (± 0.89)** | **4.27 (± 1.77)** | **0.26 (± 4.23)** | **2.82 (± 1.01)** | **3.96 (± 1.44)** | **0.23 (± 3.94)** | **2.79 (± 0.90)** |

respectively. The MAE ranges of [1.88, 4.28] mmHg, [1.28, 4.23] mmHg, and [1.35, 3.86] mmHg were reported for SBP, DBP, and MAP, respectively. These result in average (± std.) MAEs of 3.45 (± 0.89), 2.82 (± 1.01), and 2.79 (± 0.90) mmHg for SBP, DBP, and MAP, respectively, which are in line with the standard MAE of 5 mmHg defined by AAMI.

#### 2.4.4.2 Impact of Body Mass Index

The performance of mechanical sensors worn on the skin can be affected by BMI based on the coupling between the sensor and subcutaneous fat tissues. This section characterizes the impact of various BMI categories including underweight (BMI < 18.5; 1 subject), normal (18.5 < BMI < 25; 15 subjects), overweight (25 < BMI < 30; 12 subjects), and obese (BMI > 30; 4 subjects) on the accuracy of the SPG-driven monitor for BP estimation The evaluation was conducted on each group separately, as summarized in Table 2.21. The weakest performances for SBP were achieved for underweight and obese groups with MAEs of 7.52 mmHg and 6.50 mmHg, respectively, implying poorer coupling compared to normal and overweight individuals. For DBP estimation however, the lowest MAEs were achieved for normal (4.27 mmHg) and underweight (5.08 mmHg) groups. A

Table 2.21 The Impact of Blood Pressure Level on Estimation Performance

| **BMI Category (segments)** | **SBP** | | | **DBP** | | | **MAP** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **RMSE** | **ME ± SD** | **MAE** | **RMSE** | **ME ± SD** | **MAE** | **RMSE** | **ME ± SD** | **MAE** |
| Underweight (106; 3%) | 10.60 | -0.53 ± 10.64 | 7.52 | 5.55 | 0.54 ± 5.55 | 4.27 | 7.33 | 0.65 ± 7.33 | 5.49 |
| Normal (1,592; 47%) | 8.46 | -0.19 ± 8.46 | 5.64 | 7.86 | 0.26 ± 7.85 | 5.08 | 8.03 | 0.23 ± 8.03 | 5.34 |
| Overweight (1,254; 38%) | 5.61 | 0.01 ± 5.62 | 3.89 | 3.9`7 | 0.06 ± 3.97 | 2.94 | 4.30 | 0.18 ± 4.31 | 3.13 |
| Obese (401; 12%) | 10.29 | -0.29 ± 10.30 | 6.50 | 5.82 | -0.10 ± 5.83 | 2.64 | 6.26 | -0.44 ± 6.25 | 3.65 |
| **Average (± std.)** | **8.74 (± 1.98)** | **-0.25 (± 8.75)** | **5.88 (± 1.33)** | **5.80 (± 1.38)** | **0.19 (± 5.79)** | **3.73 (± 0.99)** | **6.48 (± 1.40)** | **0.15 (± 6.47)** | **4.40 (± 1.03)** |

similar pattern is observed for the estimation of MAP in which smaller BMIs resulted in higher MAEs (5.49 mmHg and 5.34 mmHg for underweight and normal groups, respectively). The contradictory observations in the estimation of SBP and DBP may indicate the influence of different data sizes across the BMI groups as demonstrated in Table 2.21. This finding necessitates further investigation using larger datasets to explore and validate its implications.

#### 2.4.4.3 Ablation Study on Feature Space

According to the analyses, the median absolute errors for SBP, DBP, and MAP are 3.81, 2.50, and 2.82 mmHg, respectively, when all feature categories are used. This section conducts an ablation study wherein the extent to which the absence of a specific feature category influences the estimation of blood pressure is investigated. For this purpose, one feature category is withheld while all others are held for training the predictive model. Fig. 2.34 quantifies the range of absolute errors for the ablated feature space for the estimation of SBP, DBP, and MAP. As can be observed, the ablation of timing, amplitude, statistical,

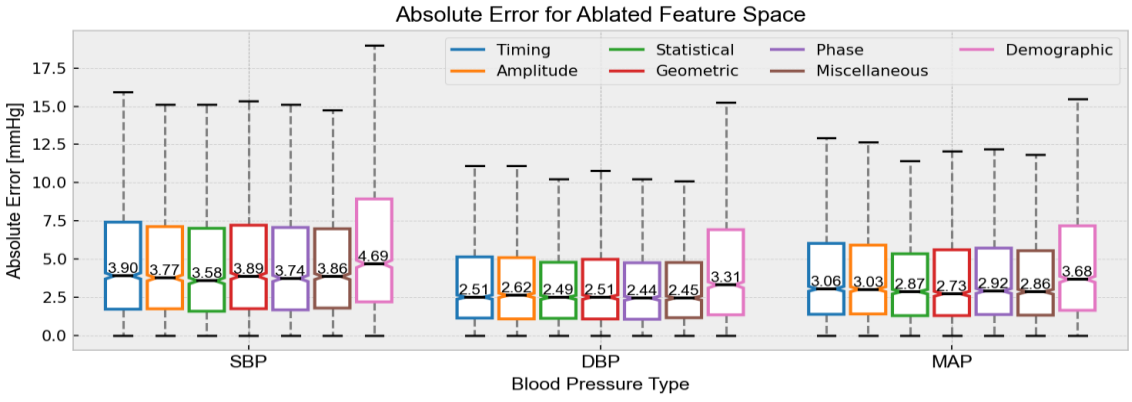


Fig. 2.34 Mean absolute error for the estimated SBP (left), DBP (middle), and MAP (right) values when feature categories are ablated.

geometric, phase, miscellaneous, and demographic feature categories introduce median absolute errors of 3.90, 3.77, 3.58, 3.89, 3.74, 3.86, and 4.69 mmHg for the estimation of SBP, respectively, which all exceed the 3.81 mmHg achieved for the non-ablated feature space. This finding demonstrates that all feature categories provide complementary information for the estimation of SBP even if some may not be listed among the top features. The same phenomenon is observed for the estimation of MAP except for the geometric features-ablated case where the median absolute error diminishes from 2.82 to 2.73. As for the estimated DBP values, the median errors decrease from 2.50 to 2.49, 2.44, and 2.45 when statistical, phase, and miscellaneous categories are ablated respectively. Fig. 2.34 witnesses the high contributions of demographic information for the estimation of SBP, DBP, and MAP by improvement margins of 0.88 (4.69-3.81), 0.81 (3.31-2.50), and 0.86 (3.68-2.82) mmHg, respectively. This evidence further supports the appearance of demographic features among the most contributing features as discussed earlier. Additionally, it is also concluded that not all the geometric features contribute to the estimation of blood pressure. Rather, only “T1_Angle1”, “T2_Angle1”, and “T7_Angle2” in combination with other feature categories could predict blood pressure level.

#### 2.4.4.4 Comparative Analysis

This section compares the performance of the proposed fingertip SPG-driven method with state-of-the-art approaches based on SPG and PPG for BP estimation. As summarized in Table 2.22, PPG-based methods employed a wide range of estimators including long short-term memory (LSTM) [178], support vector regressor (SVR) [179], ResNet [182], multi-layer perceptron (MLP) [183], sequence-to-sequence (Seq2Seq)

[184], adaptive boosting (AdaBoost) [180], and ridge regression [181]. Among PPG-based methods, the best performance for BP estimation belongs to the MLP on the MIMIC dataset with MAEs of 2.13 mmHg and 3.52 mmHg for DBP and SBP respectively, which are slightly higher than our MAEs for DBP (3.88 mmHg) and SBP (5.45 mmHg), respectively. The MLP method uses a combination of PPG and ECG sensors, making it a relatively cumbersome setting. Our sensory platform on the other hand, leverages only a compact MEMS SPG sensor at the fingertip. The SVR method trained on the University of Queensland Vital Signs (UQVS) dataset indicates DBP and SBP MAEs of 4.68 mmHg and 4.70 mmHg, respectively, suggesting a slightly superior performance to the proposed method. It is to be noted however, that our method leverages a small signal segment of 3 seconds whereas the SVR method estimates BP values based on 10-second PPG segments. This implies that our proposed solution has higher potential for wearable settings.

Table 2.22 Comparative Analysis of BP Estimation with State-of-the-Art PPG and SPG-based Methods

| **Method (year)** | **Dataset (type)** | **Input type** | **Signal length (s)** | **Performance** | |
|---|---|---|---|---|---|
| | | | | **DBP** | **SBP** |
| LSTM (2021) [178] | PPGBP (PPG) | Time, frequency, and statistical features | 30 | RMSE: 6.83, MAE: 4.68 | RMSE: 6.73, MAE: 4.70 |
| SVR (2016) [179] | UQVS (PPG) | Timing and amplitude features | 10 | MAE: 3.42 | MAE: 4.76 |
| AdaBoost (2019) [180] | MIMIC (PPG) | PPG segments | Cardiac cycle | RMSE: 4.18, ME: 0.18, and MAE: 2.43 | RMSE: 8.90, ME: -0.05, and MAE: 3.97 |
| Ridge Regression (2021) [181] | MIMIC (PPG) | PPG visibility graphs | $< 2$ | RMSE: 5.36, ME: -0.04, and MAE: 3.66 | RMSE: 8.46, ME: -0.00, and MAE: 6.17 |
| ResNet (2019) [182] | MIMIC (PPG) | PPG, PPG', and PPG" segments | 5 | MAE: 6.88 | MAE: 9.43 |
| MLP (2022) [183] | MIMIC (PPG & ECG) | PPG and ECG segments | 10 | RMSE: 3.13, ME: -0.59, and MAE: 2.13 | RMSE: 5.10, ME: -0.38, and MAE: 3.52 |
| Seq2Seq (2021) [184] | MIMIC (PPG) | PPG segments | 15 | RMSE: 7.40 and MAE: 5.56 | RMSE: 15.96 and MAE: 12.08 |
| Pulse Transit Time (2018) [187] | Proprietary (SPG & PPG) | Radial SPG and fingertip PPG segments | Cardiac cycle | MAE: 3.71 | MAE: 5.44 |
| Polyimide Film Strain Sensor (2016) [188] | Proprietary (SPG) | Radial SPG segments | 5 | ME: 1.6 | ME: -2.0 |
| **XGBoost (proposed)** | **Proprietary (MEMS Fingertip SPG)** | **Timing, amplitude, statistical, geometric, phase, miscellaneous, and demographic** | **3** | **RMSE: 6.12, ME: -0.08, and MAE: 3.88** | **RMSE: 8.03, ME: -0.02, and MAE: 5.45** |

The proposed method is also compared with a Polyimide film strain sensor [188] and pulse transit time [187] which suggest BP estimation using the SPG data of the radial artery and the combination of radial SPG and fingertip PPG, respectively. The PTT method offers a performance similar to our method for DBP (3.71 vs. 3.88) and SBP (5.44 vs. 5.45). Our proposed method outperforms the method proposed by the Polyimide film sensor in terms of ME for DBP (1.6 vs. -0.08) and SBP (-2.0 vs. -0.02).

# 3 Hemodynamic Monitoring Using Wearable Sensors

## 3.1 Mean Pressure Gradient Prediction Based on Chest Angular Movements and Heart Rate Variability Parameters

Echocardiography is the main method to grade AS severity. Mean pressure gradient (MPG) is one of the echocardiography parameters that are concordant with the severity level of the disease, and is defined as the difference in blood pressure between the left ventricle and the aorta [195]. Yet, echocardiography monitoring is expensive and limited to use in the clinic. On the other hand, wearable sensor technologies, such as electrocardiography (ECG), offer non-invasive and low-cost measurements of biological signals. Cardio-mechanical signals, representing the linear and angular chest movements induced by the heart activity, can be measured using micro-electromechanical system (MEMS) accelerometers and gyroscopes, respectively. The linear and the angular vibrations are called seismo-cardiography (SCG) and gyro-cardiography (GCG) signals, respectively. These modalities have been used for diagnosing various types of CVD in the literature [196,197].

Our research group has previously conducted studies on the detection of CVDs using SCG and GCG signals. For instance in [198], general heart rate abnormality was discussed through the morphological changes occurring in cardio-mechanical signals, where an accuracy of 99.5% was reported. Additionally, we employed SCG/GCG signals to diagnose aortic stenosis using time-frequency characteristics of the signals, where a random forest classifier predicted the existence of AS by 97.43% of accuracy [71]. Furthermore, adopting continuous wavelet transform and deep learning algorithms, a classification framework was proposed for AS detection based on multi-dimensional SCG

and GCG signals, which was capable of diagnosing AS with 98.00% of F1-score [72]. Furthermore, patients were classified 96.00% correctly through the SCG/GCG signals acquired from their chest movements into three types of CVD, with which they had been diagnosed.

In this study, we propose a wearable sensor-based method for classifying the severity of aortic stenosis. This method involves chest angular movements recorded by gyroscopes as well as heart rate variability (HRV) parameters calculated from peak-to-peak intervals of the GCG signal. As mentioned earlier, the measurement of the mean pressure gradient (MPG) across the aortic valve provides significant information about the severity of aortic stenosis [195]. On the other hand, cardiac timing intervals and HRV parameters have been shown to change with abnormalities in the cardiac activity [88]. Hence, this study targets a meaningful connection between MPG readings from echocardiography and the extracted features from GCG. To the best of our knowledge, this is the first study addressing the classification of severity level in AS through machine learning (ML) methodologies.

### 3.1.1 Methodology for Classification of Mean Pressure Gradient

#### 3.1.1.1 Experimental Setup and Data Acquisition

This study includes thirty-two AS patients (sixteen males and sixteen females). The average (standard deviation) age of the patients is 84.18 (9.61) years, where eleven, twelve, and nine patients are respectively diagnosed with mild (MPG < 30 mm Hg), moderate (30 < MPG < 50 mm Hg), and severe AS (MPG > 50 mm Hg), respectively.

Angular vibrations of the chest wall were recorded using a commercial wearable sensor node (Shimmer3 from Shimmer sensing) secured by a band strap on the mid-sternum along

the third rib. A three-axis gyroscope records the GCG signal in three dimensions. The X, Y, and Z axes correspond to the shoulder-to-shoulder, head-to-toe, and dorso-ventral directions, respectively. In this study, the dimensional letters X, Y, and Z appended as sub-scripts to GCG will denote the signal from the corresponding axis, respectively. Simultaneously, a four-lead ECG sensor was used as the reference to record the electrical activity in the heart. All waveforms were recorded at a sampling rate of 256 Hz. Immediately after, the heart rhythm and valve parameters such as MPG were also measured by an ultrasound echocardiography machine.

All data were collected at the cardiac care unit of the Columbia University Medical Center (CUMC). The subjects were seated at rest on a bed for at least five minutes. They breathed naturally without controlling their breathing depths. The patient experimental protocol was approved by the Institutional Review Board of CUMC under protocol number AAAR4104. The collected data were transferred to a computer and processed in a Python framework. The step-by-step flow graph of the developed framework for pre-processing, feature extraction, and classification is illustrated in Fig. 3.1.

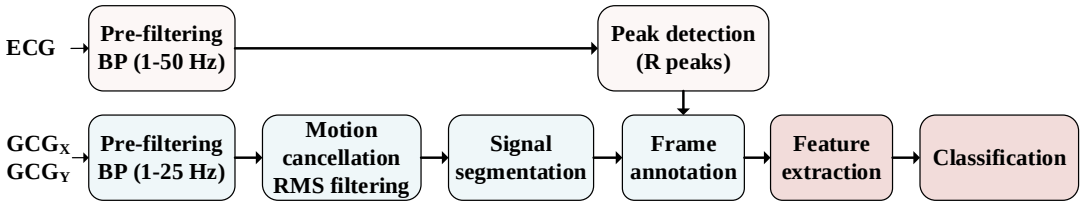


Fig. 3.1 Signal processing, feature extraction, and classification flow graph proposed in this work.

#### 3.1.1.2 Signal Pre-Processing

Firstly, to remove baseline wandering and noise artifacts, GCGX and GCGY signals were band-pass filtered using a 4th-order Butterworth filter over the range of 1-20 Hz, as shown in Fig. 3.1. The simultaneously-recorded ECG channels were band-pass filtered retaining the frequency components within the range of 1-50 Hz. Subsequently, motion artifacts associated with movements during recordings were removed from GCG signals by applying a root-mean-square (RMS) filter with a sliding window of 500 ms for signal segmentation. The segment removal threshold was selected as twice the median value of the filter. It should be mentioned that after motion artifact removal, the remaining segments were attached to each other only if there was no discontinuity between the consecutive 500-ms segments. The resulting large segments were segmented into 10-second time frames with 80% overlap between consecutive frames. For every 10-second frame, the R-peaks in the ECG signal were detected by the Pan-Tompkin algorithm [90]. The GCG frames were then annotated with I, J, K, and L points, using the R-peaks according to the methods proposed in the literature [84]. An example of the annotation is depicted in Fig. 3.2.

#### 3.1.1.3 Feature Extraction

The fiducial points on GCG signals are expected to provide valuable information about the function of the cardiovascular system [84]. Once GCG signals were annotated, two types of features were extracted from each frame: time-domain HRV parameters and GCG timing intervals.

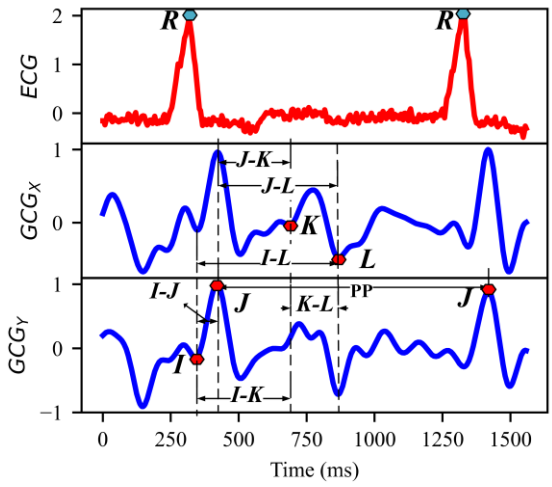


Fig. 3.2 Fiducial points of the measured signals. From top to bottom: ECG, $GCG_X$, and $GCG_Y$.

**Time-domain HRV parameters:** Time-domain HRV parameters provide valuable information about cardiac activity [91]. For temporal HRV parameters, a few time-domain analyses were applied to the series of successive inter-beat intervals (IBIs). The normal-to-normal IBI (NN) is defined as the interval between consecutive *J* peaks in the GCG signals [92]. A few HRV features were extracted from the NN time series, such as the average (AVNN), standard deviation (SDNN), root-mean-square of successive differences (RMSSD), and proportion of the number of adjacent NN intervals whose durations differ more than 50 ms (NN50) to the total number of NNs (pNN50). It is worth mentioning that SDNN, RMSSD, and pNN50 are of great clinical importance as they allow for measuring cardiac risk, respiratory arrhythmia, and parasympathetic nervous activity [91,93]. Additionally, we added the median, skewness, kurtosis, entropy (ENN), self-entropy (SENN), and conditional entropy (CENN) values of NNs to our designed feature space. Due to the nonlinearity underlying the dynamics of HRV, we also extracted the vector

angular index (VAI), the vector length index (VLI), SD1, and SD2 out of the Poincare map - a scatter plot of NN at time t in terms of NN at time t+1 [94].

**GCG timing intervals:** A few timing interval parameters describing the cardiac system were calculated for the GCG signal. It has been demonstrated that the isovolumetric contraction time (IVCT), isovolumetric relaxation time (IVRT), and left ventricular ejection time (LVET) are correlated with $I-J$, $L-K$, and $K-J$, respectively [82]. Other parameters such as the intervals between each pair of the fiducial points depicted in Fig. 3.2 along with their mean, median, standard deviation, skewness, and kurtosis values were also extracted as auxiliary features. The logic behind such an exhaustive feature extraction is to characterize the most relevant GCG timing intervals resulting in the highest accuracy for determining the severity level of AS.

#### 3.1.1.4 Predictive Models

Predictive models which are employed in this study consist of decision tree (DT), random forest (RF), support vector machine (SVM), extreme gradient boosting (XGBoost) [97], categorical boosting (CatBoost) [199], and light gradient-boosted machine (LightGBM) [200]. These models are selected as they are expected to work properly with respect to the sample space size prepared in this work. All methods except SVM are based on decision tree. The last three classifiers exploit the gradient boosting mechanism in order to minimize the prediction error, although they differ in terms of feature splitting techniques. Another reason for choosing the above-mentioned methods is to make a comparison in terms of their performance in the context of cardio-mechanical signals.

According to the severity levels mentioned earlier, a three-class dataset is provided. The dataset is trained for all the classifiers and evaluated using a validation set. Firstly, the dataset is split into two parts, training (80.00%) and test (20.00%) datasets. Then, the training dataset is fed to the predictive models, where hyperparameters are tuned in a 10-fold cross-validation (10-CV) practice. In the end, the trained models are evaluated against the test dataset to assess the robustness and generalizability of the trained models on an unseen dataset.

### 3.1.2 Experimental Results on MPG Classification

In this section, the details of the experiments are comprehensively discussed. A total of 2,878 signal frames were divided into two parts; 2,302 frames (equivalent to 25 subjects) were used for training, and the remaining 576 frames (equivalent to 7 subjects) constituted the test set. At every fold of 10-CV, the model is trained on 2,072 frames, and tested on the remaining 230 frames to optimize the hyper-parameters. In the remainder of the study, the performance evaluation is reported on the unseen dataset, i.e., the test dataset.

#### 3.1.2.1 Performance Evaluation

In this section, the performance of the proposed framework is evaluated by applying the models on the test dataset. Next, by comparing the predicted values and the true labels, the performance of the methods is reported using standard metrics such as accuracy, precision, recall, and F1-score. Table 3.1 summarizes the performance results of the predictive models introduced in the previous section. The best performance is reported by 94.44% accuracy, 94.29% F1-score, 94.45% precision, and 94.17% recall for LightGBM. The XGBoost is the runner-up predictive model that could predict the severity level by

Table 3.1 Performance Evaluation of the ML Techniques Used in This Work

| Classifier | Performance (%) | | | |
|---|---|---|---|---|
| | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| **DT** | 58.68 | 73.26 | 55.21 | 52.77 |
| **RF** | 75.17 | 78.19 | 73.86 | 74.98 |
| **SVM** | 83.33 | 83.64 | 82.44 | 82.85 |
| **XGBoost** | 93.75 | 93.80 | 93.51 | 93.64 |
| **CatBoost** | 93.75 | 93.71 | 93.44 | 93.57 |
| **LightGBM** | **94.44** | **94.45** | **94.17** | **94.29** |

93.75% accuracy, 93.71% F1-score, 93.64% precision, and 93.51% recall, quite close values to LightGBM. As such, the proposed framework is indicated to be sufficiently capable of classifying the severity level of stenosis from GCG readings. As expected, gradient-based boosting methods outperform other simpler tree-based algorithms such as DT and RF, whereas SVM leveraging a hinge loss could predict the severity level by 83.33% accuracy. This indicates the superiority of state-of-the-art tree-based methods over SVM. Furthermore, CatBoost shows almost the same performance as XGBoost in spite of its higher complexity in the feature splitting procedure. Having analyzed the reported performance, it is concluded that the proposed feature space highly correlates with the severity level of aortic stenosis, which is a promising achievement.

#### 3.1.2.2 Top Features

A significant point that discriminates the tree-based methods from other machine learning techniques is their interpretability in terms of performance. This means that the top features, best representing the outcome, could be determined using feature importance scores. The f-scores provided by the gradient-based models consider the impacts of the entire samples on the output. In this work, however, we found the top features using the

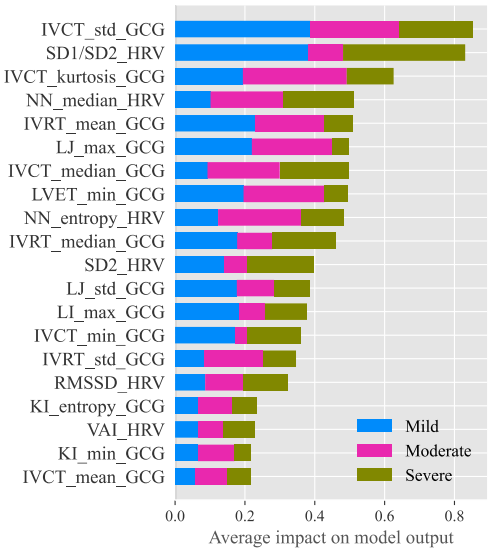


Fig. 3.3 Top features for AS severity classification and their average impact on the class score ranked by Light GBM.

Shapley Additive exPlanations (SHAP) technique, a game theory-based approach for interpreting the output of a model in terms of the input feature space. SHAP calculates the feature importance in terms of the impact of every single observation on the output [105,201]. Fig. 3.3 demonstrates the average impact of the top 20 features on the model output. As illustrated, six features from HRV parameters and fourteen features from GCG timing intervals constitute the top features. IVCT and IVRT that correspond to ventricles iso-volumetrically contraction time and the time interval between aortic valve closure and the onset of the opening of mitral valve, respectively, hold the highest frequencies among the top GCG intervals. Among the top HRV features, three are listed from the Poincare characteristic, whereas the remaining are from the typical HRV features calculated from NN. Considering the points mentioned above, IVCT, IVRT, and Poincare HRV features are the most correlated features with the severity level of AS.

## 3.2 Detection of Left Ventricular Ejection Fraction Using Fusion of Accelerometer Contact Microphone and Biopotential Sensors

Heart failure (HF), defined as a condition in which the cardiac muscle fails to pump adequate blood to meet the needs of the body, affects 6.2 million in the US annually [8]. HF is primarily accompanied by gradual decreases in the volumetric fraction of the blood ejected from the left ventricle per heartbeat [202]. This fraction is referred to as left ventricular ejection fraction (LVEF) which normally falls within the range of 52%-72% [203]. Lower percentages of LVEF suggest cardiovascular abnormalities such as cardiomyopathy, heart valve problems, and chronic hypertension [204], which require timely medical intervention. As such, continuous monitoring of LVEF in patients with cardiovascular diseases is critical, contributing to timely risk management. This motivates the need for round-the-clock monitoring devices.

LVEF is predominantly monitored by echocardiogram (echo) machines in the clinic. Echo utilizes ultrasound technology to produce images of the heart, and accurately quantify the blood volume ejected from the left ventricle at each heartbeat. Nevertheless, the bulky nature of echo monitors inhibits their use as ubiquitous monitoring devices. Several studies have addressed cardiovascular monitoring using phonocardiogram recorded by a stethoscope [205–207]. While stethoscopes provide useful information about mechanical activities of the cardiac system, their use for real-world continuous monitoring is limited due to susceptibility to ambient noise. On the other hand as demonstrated in [208] and [128], accelerometer contact microphones are capable of monitoring the cardiopulmonary system and pulmonary disorders respectively.

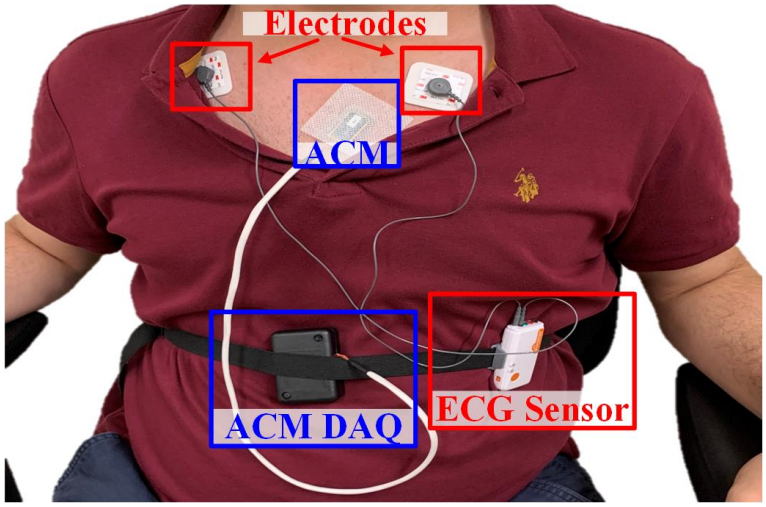


Fig. 3.4 The experimental setup including an accelerometer contact microphone (ACM from StethX Microsystems) on the chest connected to a data acquistion (DAQ) module, and a single lead of an electrocardiogram (ECG) sensor node (Shimmer3 ECG Development Kit) strapped around the torso.

The authors in [209] have demonstrated that LVEF is proportional to cardiac time intervals extracted from heart sounds. In this study, we revisit the detection of left ventricular ejection fraction, and adapt it to wearable settings. To this end, we establish a monitoring platform based on the fusion of acoustic and biopotential characteristics of the precordium recorded by the recently-developed accelerometer contact microphone (ACM) in [128] and an electrocardiogram (ECG) sensor respectively. This framework is comprehensively discussed in the following sections.

### 3.2.1 Experimental Setup & Methodology

#### 3.2.1.1 Data Collection

This study includes seven cardiovascular patients (5 males and 2 females) with average (± standard deviation) LVEF of 27.85% (± 12.86%). The average (± standard deviation) age, height, and weight of the patients are 72 (± 13.97) years, 170.18 (± 12.44) cm, and 73.03 (± 20.79) kg, respectively. Additionally, eight subjects (3 males and 5

females) with average (± standard deviation) LVEF of 58.75% (± 2.31%) constitute the healthy cohort whose average (± standard deviation) age, height, and weight are 60 (± 18.05) years, 169.22 (± 11.83) cm, and 85.21 (± 18.02) kg, respectively. All subjects are studied at the cardiac care unit of Sorin Medical P.C. under measurement protocols approved by the Institutional Review Board (IRB) of Stevens Institute of Technology with protocol number 2022-044 (N).

Fig. 3.4 describes the experimental setup designed for LVEF abnormality detection. A $\pm 4$ g sensitive accelerometer contact microphone (ACM) with micro-g resolution and a small form-factor of 27 mm × 15 mm × 2.5 mm from StethX Microsystems is attached to the pulmonary region of the subjects using medical-grade adhesive tape. This device is a low-noise accelerometer with a wide operational bandwidth of 0-10 kHz, allowing for recording heartbeat-induced sounds and vibrations on the chest wall. The ACM is not sensitive to airborne emission sounds, making it a robust phonocardiogram sensor against acoustic ambient noise. The subjects were seated at rest on a chair for a period of five minutes, followed by five minutes of ACM measurements at a sampling rate of 22.33 kHz, still at a seated position. Simultaneously, the left arm-right arm (LA-RA) lead of a four-lead ECG sensor (Shimmer3 ECG Development Kit; Shimmer, Dublin, Ireland) was used to record the electrical activities of the cardiac system at a sampling rate of 512 Hz. The patients' LVEF values were also measured by an ultrasound echocardiography machine as the reference modality. The collected data were transferred to a computer and processed in a Python framework.

### 3.2.1.2 Signal Processing for PEP/LVET Estimation

Fig. 3.5 represents the signal processing chain for LVEF abnormality detection. A zero-phase, 3rd-order Butterworth filter over the frequency range of 10 Hz - 2 kHz is initially applied to ACM recordings to mitigate noise, respiration artifacts, and baseline wander. The filtered signals are then downsampled to a sampling rate of 4 kHz to accelerate the processing steps. In order to determine the first and second heart sounds ($S_1$ and $S_2$ respectively), the envelope of the signal is calculated using a homomorphic filter [210]. This results in the appearance of peaks corresponding to the main heart sounds ($S_1$ and $S_2$) as well as murmurs and the remnant noise in the envelope. To extract $S_1$ and $S_2$ from ACM recordings, a peak detector algorithm is employed to filter out the peaks corresponding to noise and heart murmurs. This algorithm considers a peak as a heart sound only if its amplitude is greater than 0.2 times the maximum value of the envelope, and its distance from the previous heart sound is at least 100 ms. Finally, the detected peaks are labeled as $S_1$ and $S_2$. This is performed by comparing the time differences of a peak with its preceding

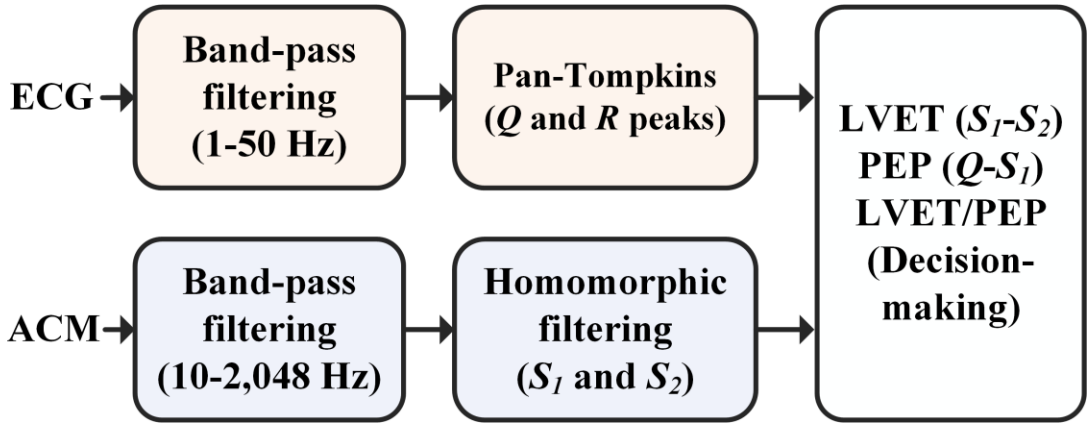


Fig. 3.5 The signal processing chain including pre-filtering, signal annotation, and decision-making stages for noise cancellation, time intervals estimation, and LVEF abnormality detection, respectively.

and following peaks. If the time difference of a peak with its following peak exceeds the time difference of that peak with its preceding peak, the peak is labeled as $S_2$; otherwise, it is considered $S_1$.

As shown in Fig. 3.5, the simultaneously-recorded ECG signal is band-pass filtered retaining the frequency components within the range of 1-50 Hz. Furthermore, to attenuate the power-line interference, a second-order infinite impulse response (IIR) notch filter centered at 60 Hz with a quality factor of 25 is applied to the filtered ECG signal in forward and backward directions. The *R*-peaks in the ECG signal are detected by the Pan-Tompkin algorithm [90]. The *Q*-peaks are then annotated by finding the minimum value of the signal within 50 ms backward of the corresponding *R*-peaks.

After annotation of ACM and ECG signals, the left ventricular ejection time (LVET), which is defined as the time interval between aortic valve opening ($S_1$ in ACM

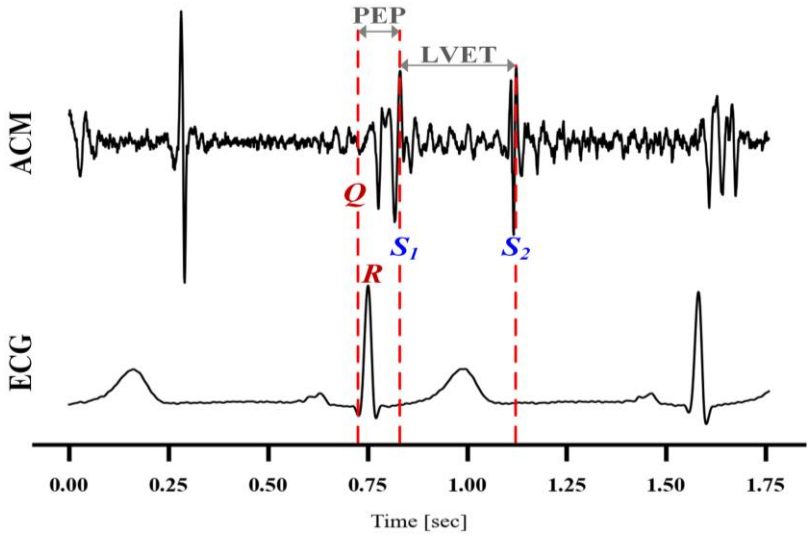


Fig. 3.6 ACM (top) and ECG (bottom) annotations and time intervals, including pre-ejection period (PEP) and left ventricular ejection time (LVET) corresponding to time difference of Q-S1 and S1-S2, respectively.

signals) and its closure ($S_2$ in ACM signals), is estimated per heartbeat as shown in Fig. 3.6. Additionally, the pre-ejection period (PEP), which signifies the time elapsed between the electrical depolarization of the left ventricle (Q peak in ECG) and the beginning of the ventricular ejection ($S_1$ in ACM), is estimated. It has been demonstrated that LVEF is inversely proportional to PEP/LVET [209]. In this study, we leverage a fusion of ACM and ECG sensory data to detect LVEF abnormality. The use of ACM in this study is significant. The aortic closure point in accelerometric data is often contaminated by noise, hindering the accurate estimation of LVET. Owing to its state-of-the-art sensitivity and low noise, the ACM on the other hand provides $S_2$ sounds which correspond to the aortic closure, but with more evident patterns in the signal. This helps estimate LVET periods with much higher confidence. As shown in Fig. 4.6, heart sounds are represented by evident peaks resulting from the absolute robustness of the ACM against the ambient noise. *This allows for wearable monitoring of LVEF, which traditionally was not feasible with hand-held stethoscopes due to their susceptibility to background noise and low signal-to-noise ratios (SNRs).*

### 3.2.2 Experimental Results and Discussion

#### 3.2.2.1 LVEF Abnormality Detection Performance

Table 3.2 summarizes the performance of the proposed method for LVEF abnormality detection for 8 healthy subjects and 7 abnormal LVEF patients who are labeled by H and P respectively. Per subject, the estimated PEP/LVET value for each beat is compared with the standard range (i.e. 0.3-0.4 as defined by [211]) for healthy subjects to separate healthy and abnormal subjects. If the estimated PEP/LVET value falls within 0.3-

0.4, it is concluded that the subject is healthy; otherwise, the subject is classified with abnormal LVEF. If the label of the subject is consistent with the predicted class, the beat is called a true beat; otherwise, it is called a false beat. As such, the number of true and false beats in Table 3.2 respectively represent truly and falsely classified heartbeats with respect to the ground-truth labels (H or P). An average (± standard deviation) accuracy of 84.47% (± 17.58%) is achieved for 5,989 heartbeats corresponding to all subjects together, suggesting the accurate performance of the proposed method. An average (± standard deviation) accuracy of 84.06% (± 15.86%) is obtained for healthy subjects where the highest and lowest classification accuracies among healthy subjects are reported for subject 1 with 100% and subject 4 with 57.09%, respectively. It seems that the ACM recording associated with subject 4 is distorted in a considerable number of segments due to subject movements, resulting in a low SNR and 136 false beats. On the other hand, no distortion or motion artifacts are observed in the ACM recordings of subject 1, leading all 467 beats to be correctly classified.

As reported for subjects with abnormal LVEF values (subjects 9-15 in Table 3.2), accuracies of 100% are achieved for subjects 10, 12, and 14, whereas the largest classification error is reported for subject 11 with accuracy and falsely classified number of beats of 54.35% and 194 respectively. An average (± standard deviation) accuracy of 84.91% (± 20.66%) signifies the classification performance for subjects with abnormal LVEF. Comparing the average accuracies obtained for the healthy cohort and abnormal subjects (84.06% vs. 84.91%), no statistically significant difference is observed between the two groups ($p < 0.05$), implying a robust performance for a variety of LVEF values.

#### 3.2.2.2 Limitations

Although the proposed method is capable of an adequate level abnormal LVEF detection (84.47%), it still misses almost 15% of abnormalities. This limitation can be addressed by extracting additional time intervals rather than relying merely on LVET and PEP, similar to our previous studies on valvular heart diseases (VHDs) [125,126]. In this study, a decision-making algorithm based on the normal range of PEP/LVET is used. However, LVEF abnormality detection can benefit from machine learning techniques encouraging data-driven modeling for abnormality identification.

Another limitation of this work is the small number of subjects participating in the study. The larger the number of subjects, the more the number of left ventricular ejection

Table 3.2 Performance Evaluation of Abnormality Detection (0.3 < PEP/LVET < 0.4)

| Subject | # Beats | # True beats | #False beats | Accuracy (%) |
|---|---|---|---|---|
| **1 (H)** | 467 | 467 | 0 | 100 |
| **2 (H)** | 531 | 426 | 105 | 80.22 |
| **3 (H)** | 510 | 496 | 14 | 97.25 |
| **4 (H)** | 317 | 181 | 136 | 57.09 |
| **5 (H)** | 473 | 449 | 24 | 94.92 |
| **6 (H)** | 335 | 329 | 6 | 98.21 |
| **7 (H)** | 553 | 406 | 147 | 73.41 |
| **8 (H)** | 557 | 398 | 159 | 71.45 |
| **9 (P)** | 424 | 422 | 2 | 99.52 |
| **10 (P)** | 364 | 364 | 0 | 100 |
| **11 (P)** | 425 | 231 | 194 | 54.35 |
| **12 (P)** | 371 | 371 | 0 | 100 |
| **13 (P)** | 328 | 271 | 57 | 82.62 |
| **14 (P)** | 334 | 334 | 0 | 100 |
| **15 (P)** | 409 | 237 | 172 | 57.94 |
| **Average (± std.)** | **426.53 (± 84.79)** | **358.8 (± 93.77)** | **67.73 (± 75.05)** | **84.47 (± 17.58)** |

fraction values, resulting in more generalizable models. Furthermore, the average age of the subject population is around 65 years which is not representative of LVEF abnormalities at younger ages. As such, extensive data collection could benefit our modeling.

## 3.3 Fingertip Strain Plethysmography: Representation of Pulse Information based on Vascular Vibration

Plethysmography is a non-invasive technique for the evaluation of blood dynamics in peripheral vessels. Over the past few decades, plethysmography has demonstrated promise for the estimation of physiological parameters such as blood oxygen saturation [212], respiratory rate [213], heart rate variability [214], and blood pressure [215]. Such a wide range of clinical utilities has promoted the emergence of various types of plethysmography methods as described below.

Photoplethysmography (PPG), the most common plethysmography method, is an optical sensing strategy in which a light-emitting diode is coupled with a photodetector. The light source emits green, red, or infrared light, and the photodetector receives the light reflected from the vessel. Under the Beer-Lambert-Bouguer law, the incidence of light on a volume of blood causes part of the light energy to be absorbed by blood components. As a result, the reflected light has lower density compared to the transmitted light [61]. Hence, temporal changes in blood volume due to heartbeat appear as a pulsatile pattern in the PPG signal. It has been demonstrated that dark skin color and motion artifacts could impose significant noise in PPG recordings [216].

Impedance plethysmography (IPG), also referred to as rheography, is another type of plethysmography which is based on the measurement of the electrical impedance of the body part. Blood has a lower impedance against electrical current in comparison to skin, muscles, and bones. As the blood flows along a body segment such as the arm, the impedance of blood becomes parallel with the impedance of other tissues, resulting in a decrease in total impedance. As such, continuous changes in the impedance represent the cardiac cycles [217]. The IPG setup consists of 4 electrodes placed along a line on a body segment where the two outer electrodes measure the current and the two inner electrodes pick up the voltage. Thereby, the impedance is reported based on the Ohm's law.

Another method is strain gauge plethysmography (SPG) which involves a fine-bore rubber tube filled with mercury placed around the limb. Following blood flow variations during each cardiac cycle, the circumference of the tube changes, and thus the impedance of mercury indicates a heartbeat signal [218]. Despite excellent signal qualities, the uses of IPG and SPG in clinical settings are limited due to their cumbersome setup.

In this study, we introduce a fingertip strain plethysmography, which can be measured with a small sensor and has high robustness against optical limitations. *The proposed approach leverages a sensitive MEMS strain sensor to directly sense the vascular pulsations on the skin in a wearable setting*. The captured vibration could provide the dynamics of the cardiac cycle, allowing the proposed technology to be reliably used as a replacement or an augmentation to existing plethysmography methods.

### 3.3.1 Fingertip Strain Plethysmography Methodology

As the heart beats, the blood flowing through the circulation system exerts a force on the vascular wall at each site, resulting in a change in the diameter of the vessel [219]. This force is then transferred to the skin surface through the tissues, causing minute vibrations on the skin [218]. Hence, the morphology of the cardiac-induced vibration on the vascular wall is expected to follow the temporal pattern of the cardiac cycle.

The small form factor, low power consumption, and cost-effectiveness of micro-electromechanical systems (MEMS) sensors make them ideal candidates for the characterization of cardiovascular-induced vibrations. Seismocardiogram and ballistocardiogram, representing the movement of the chest and the whole body as a result of blood ejection respectively, are two examples of such uses. These modalities suggest great potential for various applications, such as the detection of cardiovascular diseases [44,125,126,132,220–222] and blood pressure monitoring [223]. Similarly, MEMS strain sensors have exhibited considerable capability for the monitoring of vital signs [224].

#### 3.3.1.1 Experimental Setup

Fig. 3.7 (a) demonstrates the experimental setup for the acquisition of the fingertip strain plethysmography signal in a wearable setting. In this study, the vascular pulsations were registered from the tip of the index finger on the left hand. For this purpose, we employed a custom MEMS strain sensor. The small form factor of the sensor head allows for its tight placement on the fingertip using a piece of medical-grade tape. As shown in Fig. 3.7 (a), the sensor picks up pulsation at the fingertip.

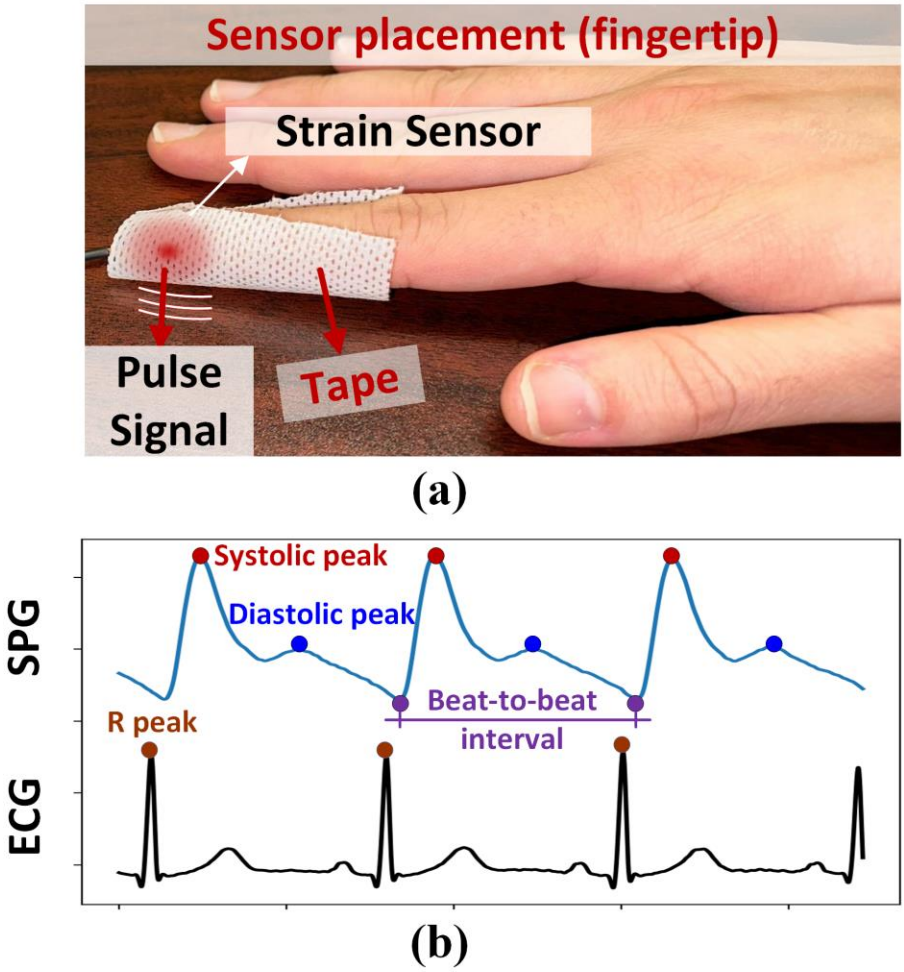


Fig. 3.7 The setup of wearable fingertip strain plethysmography. The strain sensor is placed on the fingertip to collect pulse information and hemodynamic parameters. (b) The fingertip pulsation signal and its fiducial points.

The experiment was conducted on 10 healthy subjects, consisting of nine males and one female. The average (± standard deviation) age, height, and weight of the subjects are 24.10 (± 1.79) years, 171.10 (± 11.06) cm, and 77.20 (± 13.71) kg, respectively. The experiment began with each subject stair-stepping for 5 minutes as a strategy to elevate the heart rate. Immediately afterwards, the subject was seated on a chair in a stationary, upright position for five minutes during which the strain plethysmography signal was recorded from the

fingertip. Additionally, a single-lead ECG sensor node [225] (Shimmer Sensing, Dublin, Ireland) was strapped around the chest wall with the electrodes attached to the right arm (RA), left arm (LA), and right leg (RL), to simultaneously record the RA-LA ECG signal as a reference for beat-to-beat intervals. This module also contains an inertial sensor which is used as a reference for respiratory cycles by recording torso movements occurring due to inhalation and exhalation from the chest wall. This reference, although not an ideal gold standard, will be used in this study to demonstrate proof-of-concept results. The ECG and inertial data were recorded at a sampling rate of 512 Hz. All the collected data were transferred to a computer for processing and analyses. The patient experimental protocol was approved by the Institutional Review Board of Stevens Institute of Technology under protocol number 2022-044 (N).

An example of the strain plethysmography recording along with its corresponding ECG signal is depicted in Fig. 3.7 (b). Visual inspection suggests great consistency between the SPG and ECG signals in terms of heartbeat information. In each cycle, two prominent peaks are observed, one associated with the onset of systole (marked with red circles) and the other with the onset of diastole (blue circles). The onset of each beat is the inflection point between the diastolic peak and the following systolic peak. The inflection points, which are marked by purple circles, can be used to estimate beat-to-beat intervals and heart rate values. For this purpose, the fiducial points need to be annotated in the recordings, the process of which is described in the following section.

### 3.3.1.2 SPG Signal Annotation

The high sampling rate of the SPG recordings provides a broad range of frequency components, including both low and high-frequency components corresponding to hemodynamics and noise, respectively. In this proof-of-concept study, however, we will

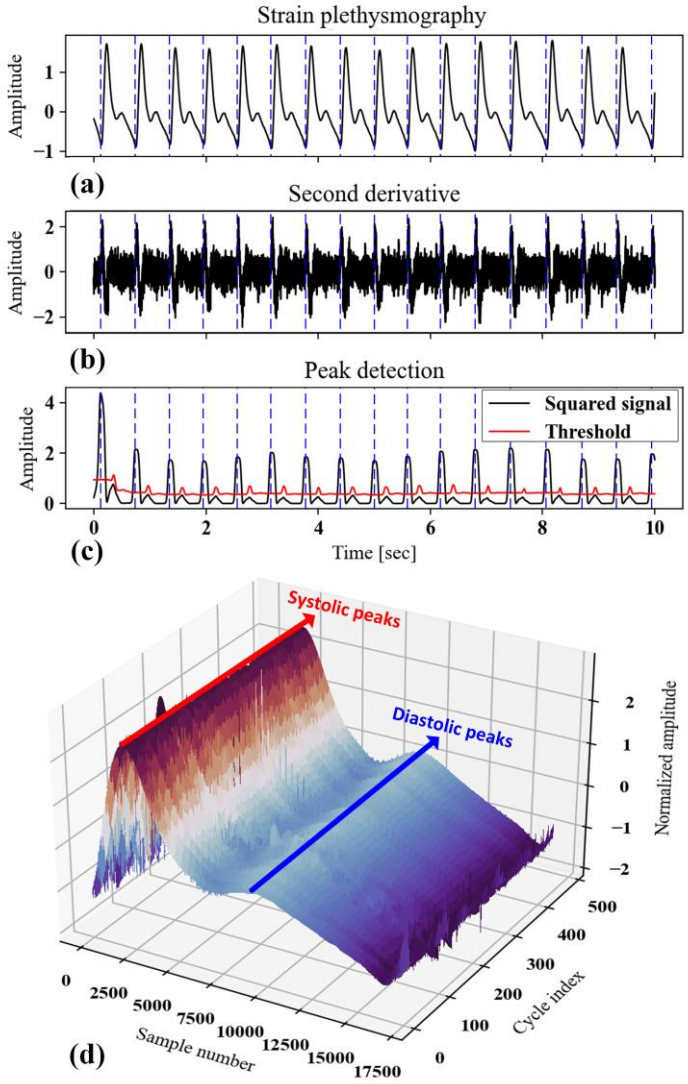


Fig. 3.8 Signal segmentation and annotation of a 10-second SPG signal. (a) The band-pass filtered SPG signal. (b) the second derivative of the signal, with peaks corresponding to the inflection points in the original SPG signal. (c) The squared version of the second-derivative and the threshold for peak detection. (d) The cardiac cycles extracted from the 5-minute SPG signal.

only investigate the low frequency components. As such, the SPG signals were first band-pass filtered within 0.1-15 Hz using a second-order Butterworth filter in forward and backward directions as shown in Fig. 3.8 (a). In order to locate the starting point of each cardiac cycle, the inflection points in the SPG recordings need to be determined. For this purpose, the second derivative of the SPG signal which provides information about the curvature direction is calculated. As illustrated in Fig. 3.8 (b), the local maximum points in the resulting signal correspond to the inflection points, i.e., the starting points in the cardiac cycle, in the original SPG signal. An adaptive thresholding method can be used to locate the local maximum points. Firstly, the second-derivative signal is clipped by discarding the segments with negative values. The clipped signal is then squared, from which a threshold value is adaptively determined using the moving-average filter as follows:

$$\mathrm{K}[n] = \beta \times \mu + \frac{1}{2M+1}\sum_{i=n-M}^{n+M} x[i]\,, \qquad (1)$$

where $x[n]$, $\mathrm{K}[n]$, $2M+1$, and $\mu$ denote the clipped signal, the threshold, the length of the moving-average filter (0.667 seconds), and the average value of the clipped signal, respectively. The mean value in this equation is normalized by $\beta$, the optimal value of which was found to be 0.02 in this study. As a result, the signal segments exceeding the threshold are kept for peak detection, and the rest of the signal is set to zero. The maximum peaks corresponding to the inflection point in the original SPG signal can be found within the remaining blocks. These points represent the onset of the cardiac cycle. Fig. 3.8 (d)

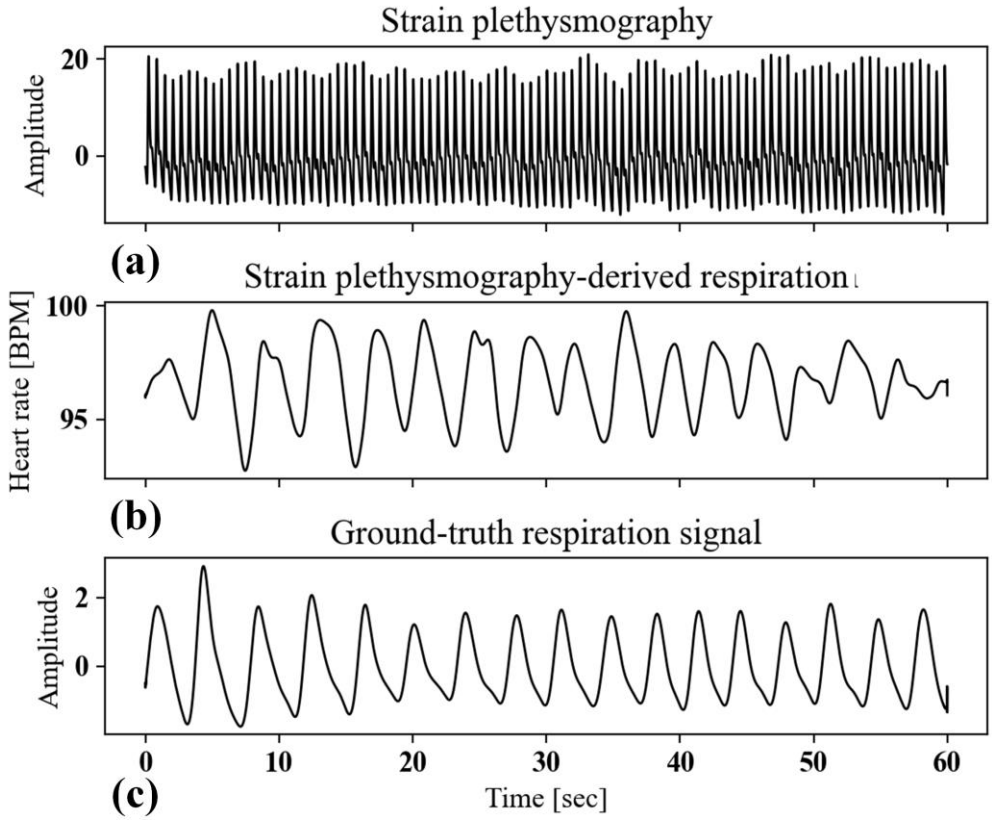


Fig. 3.9 The extraction of respiration from SPG signal. (a) A 1-minute SPG signal. (b) The SPG-derived respiration signal. (C) The reference for respiration signal from Shimmer.

illustrates a 3D representation of the 482 cardiac cycles detected in a 5-minute SPG recording in which the X, Y, and Z axes show the cycle index, sample number, and normalized amplitude respectively. For the sake of presentation, the cardiac cycles were resampled to the length of the longest heartbeat duration of the subject, i.e., 0.68 seconds. Visual inspection suggests a uniform pattern for healthy subjects where systolic and diastolic peaks are aligned with the red and blue lines respectively shown in Fig. 3.8 (d). Following the cycle segmentation, the global maximum within each cycle is annotated as

the systolic peak while the maximum after the first local minimum is marked as the diastolic peak.

#### 3.3.1.3 Hemodynamic Estimation

Each 5-minute SPG recording is firstly segmented by a 1-minute window and a step of 5 seconds between consecutive segments (i.e., 92% overlap). The following hemodynamic parameters are then extracted from each SPG segment and the reference sensor to check the reliability of the vascular vibrations for cardiovascular monitoring.

**Heart rate (HR)** is defined as the number of heart beats per minute (BPM) and provides valuable information about the presence of tachycardia and bradycardia [226]. This parameter is estimated by averaging the beat-to-beat intervals in each 1-minute SPG segment.

**Heart rate variability (HRV)** is defined as the fluctuation in the time intervals between adjacent heartbeats [91]. There are many HRV parameters in time and frequency domains. As an example, we use the standard deviation of the normal-to-normal beat intervals (SDNN) in this study.

**Respiratory rate (RR)** represents the number of respirations per minute (RPM). It has been demonstrated in the literature that heart rate varies throughout the respiratory cycle. It increases during inhalation and decreases during exhalation [227]. This well-known phenomenon occurs due to the automatic regulation of the cardiovascular momentum during respiration, which is called respiratory sinus arrhythmia (RSA) [227]. Fig. 3.9 depicts the extraction of the respiration signal from a 1-minute SPG signal. Firstly, the signal is divided into sub-segments, and the average HR is estimated in each sub-segment.

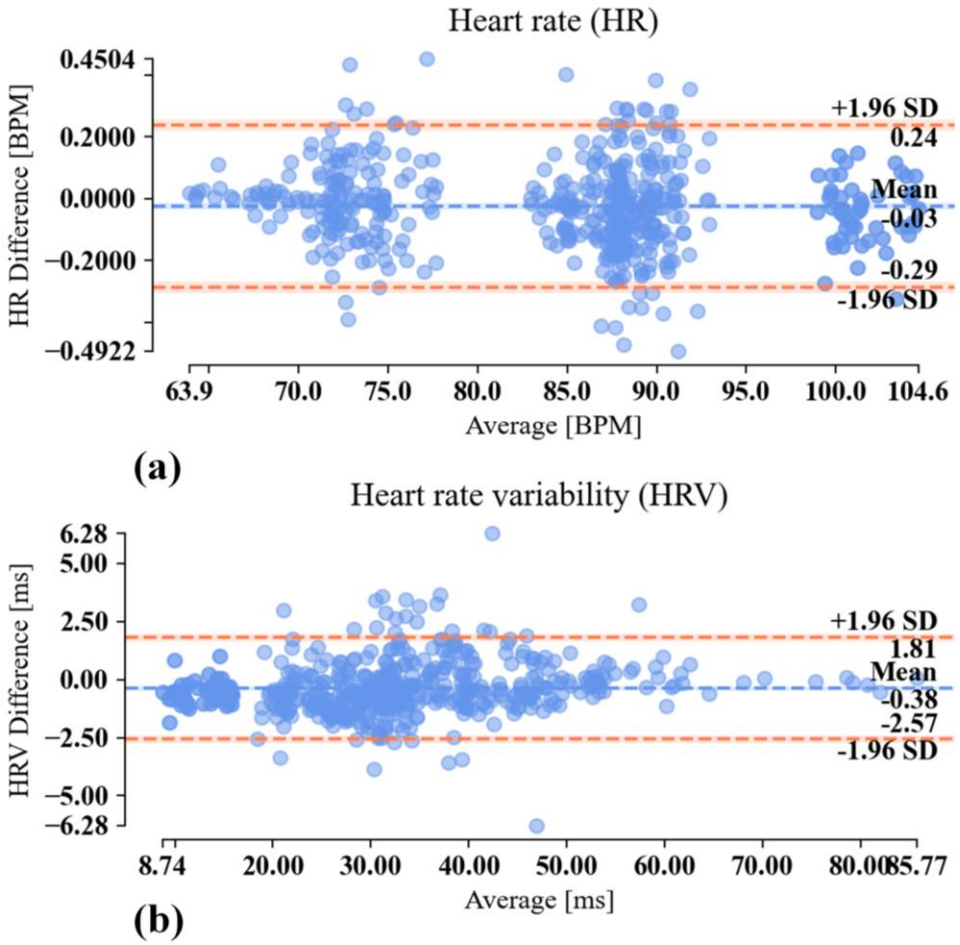


Fig. 3.10 The Bland-Altman characteristics for (a) heart rate and (b) heart rate variability, estimated by SPG and ECG.

Hence, each sub-segment should include at least one cardiac cycle. Considering the range of HR within 48-150 BPM, the sub-segments are selected as 1.25 seconds (the shortest length to ensure the existence of a valid HR in each sub-segment). In order to increase the resolution, a 70% overlap was considered between consecutive windows. Fig. 3.9 (b) illustrates the respiration signal derived from the 1-minute SPG signal shown in Fig. 3.9 (a). Comparing the reference signal from Shimmer in Fig. 3.9 (c) and the signal derived from SPG suggests excellent consistency in terms of the number of peaks and their

placements. In order to estimate the RR, we perform a fast Fourier transform (FFT) on SPG-derived and reference signals. The dominant peak represents the RR in each resulting signal, based on which the accuracy is calculated as explained in the following section.

### 3.3.2 Proof-of-Concept Results & Discussion

In this section, statistical analyses are conducted to compare HR, HRV, and RR values estimated from SPG with their ground-truth values from Shimmer. As mentioned earlier, the SPG recording of each subject is segmented into 1-minute segments. This results in a total of 555 segments across the ten subjects. Fig. 3.10 demonstrates the Bland-Altman characteristics for the HR and HRV based on 95% of agreement (1.96 × Std.) between SPG and Shimmer. As shown in Fig. 3.10 (a), the comparison is valid for a wide range of HR values from 63.9 BPM to 104.6 BPM. The mean difference between the HR values estimated by SPG and ECG is -0.03 BPM, implying a negligible estimation bias for SPG. Furthermore, 95% of the differences fall within the range of -0.29-0.24 BPM which suggests excellent consistency between the two modalities. Fig. 3.10 (b) illustrates the comparison of the mean HRV values within the range of 8.74-85.77 ms. A mean difference of -0.38 and agreement limits of -2.57 ms and 1.81 ms are reposted for SDNN which demonstrates the reliability of the SPG modality. The slight differences between the HRV values obtained from SPG and ECG, although minor, provoke the need for a meticulous investigation to find the underlying reasons.

The accuracies of the SPG-derived hemodynamics are also evaluated in this study using the following criterion:

$$\text{Accuracy} = 100 - 100 \times \left|\frac{P_{SPG} - P_{ref}}{P_{ref}}\right|, \qquad (2)$$

where $P_{SPG}$ and $P_{ref}$ denote the hemodynamic parameters estimated by SPG and the reference sensor, respectively. The accuracies for HR, HRV, and RR are summarized in Table 3.3. As reported, the average (± standard deviation) accuracy for HR of the ten subjects is 99.88% (±0.04) where the lowest accuracy (i.e., 99.83%) belongs to subjects 4 and 9 and the highest accuracy (i.e., 99.94%) to subjects 5 and 7. As for the estimated HRV values, the accuracy suggests an average (± standard deviation) of 96.43% (±1.44), demonstrating a reliable estimation of SDNN from SPG. The accuracy for HRV values across the ten subjects ranges from 94.42% (±3.75) for subject 3 (the lowest agreement) to 98.89% (±1.12) for subject 7 (the highest agreement). Due to the movement of the strap

Table 3.3 Estimation Accuracies Achieved for Hemodynamic Parameters

| Subject | Heart rate (%) | Heart rate variability (%) | Respiratory rate (%) |
|---|---|---|---|
| **1** | 99.84 (±0.12) | 95.91 (±3.35) | 95.24 (±4.76) |
| **2** | 99.84 (±0.13) | 97.25 (±2.94) | 91.34 (±3.10) |
| **3** | 99.91 (±0.06) | 94.42 (±3.75) | 90.96 (±6.56) |
| **4** | 99.83 (±0.12) | 97.38 (±2.13) | 95.70 (±3.04) |
| **5** | 99.94 (±0.04) | 96.54 (±2.02) | No reference |
| **6** | 99.91 (±0.06) | 94.42 (±3.75) | 91.95 (±6.36) |
| **7** | 99.94 (±0.05) | 98.89 (±1.12) | 89.01 (±4.14) |
| **8** | 99.87 (±0.07) | 94.69 (±2.31) | 94.34 (±4.53) |
| **9** | 99.83 (±0.10) | 97.35 (±2.37) | No reference |
| **10** | 99.84 (±0.11) | 97.44 (±2.55) | No reference |
| **Average (±std)** | **99.88 (±0.04)** | **96.43 (±1.44)** | **92.64 (±2.30)** |

holding the reference accelerometer on the chest wall, the reference respiration signal was not captured well for subjects 5, 9, and 10. Therefore, in the column corresponding to respiration accuracies, their cells are marked by “No reference”. The average (± standard deviation) estimated accuracy for the respiratory rate of the remaining 7 subjects is reported as 92.64% (± 2.30) which implies a high capability for RR estimation using SPG. The slight inconsistency between the SPG-derived RR and reference RR can be addressed using more advanced techniques for the extraction of respiration from heartbeat signals, such as a time-frequency analysis, as well as using a more reliable respiration reference sensor.

## 3.4 A Comparative Study of Heart Rate Variability Parameters Estimated from Strain Plethysmography Recordings of Radial and Finger Arteries

Heart rate variability (HRV) analysis is a valuable tool for assessing autonomic nervous system function, reflecting the balance between sympathetic and parasympathetic activities [91]. HRV biofeedback can guide individuals to consciously regulate their breathing and engage in relaxation for stress management [228]. Furthermore, HRV parameters have indicated promise in characterizing cardiovascular conditions such as arrhythmias [229], coronary artery disease [230], and aortic stenosis [125,126]. Therefore, integrating accurate estimation of HRV parameters into daily routines is essential.

In recent years, wearable devices have gained popularity for non-invasive and continuous HRV monitoring in real-world settings. Among wearable technologies, photo-plethysmography (PPG) and electrocardiography (ECG) offer reliable assessments. PPG measures blood volume changes in peripheral blood vessels, using optical sensors, and provides valuable information about heart rate and blood flow dynamics. ECG on the other

hand captures the electrical activity of the heart using electrodes placed on the skin, enabling accurate determination of R-R intervals for HRV analysis. While PPG and ECG have been successful in HRV estimation, they do have certain limitations.

The limitation of PPG is its susceptibility to motion artifacts [186] and skin tone variations [185]. Body movements and darker skin tones can introduce noise and affect the accuracy of PPG-derived HRV parameters. While ECG provides highly accurate R-R intervals for HRV analysis, it requires direct contact with the skin through electrodes, which can be cumbersome for long-term monitoring and may cause skin irritations. These limitations of existing modalities highlight a need for new approaches to HRV parameter estimation.

Strain plethysmography measures changes in arterial diameter and blood volume using strain sensors placed on the skin surface [231]. This modality has the advantage of being non-invasive, contactless, and less susceptible to motion artifacts compared to PPG and ECG. By capturing arterial dynamics, SPG can potentially provide comprehensive and accurate information regarding cardiac cycle events.

In this study, *we investigate the feasibility of estimating HRV parameters using wearable micro-electromechanical systems (MEMS) strain plethysmography from radial and digital arteries*. Furthermore, we compare the results with those obtained from PPG signals in the literature. This work contributes to the development of strain-based methods for continuous monitoring of HRV parameters and sets a precedent for future studies.

### 3.4.1 Experimental Setup and Methodology

#### 3.4.1.1 Data Collection and Preparation

Fig. 3.11 illustrates the experimental setup for acquiring the strain plethysmography signal. Our approach involved utilizing a custom-built, highly sensitive capacitive MEMS strain sensor affixed to the subject's fingertip (on the digital artery) and wrist (on the radial artery) using medical-grade adhesive tape, as shown in Fig. 3.11 (a). A wafer-level-packaged capacitive MEMS strain sensor (depicted in Fig. 3.11 (b)) was developed for recording the SPG signal. This sensor utilizes a parallel plate capacitor with a gap size of 200 nm, enabling an extremely high sensitivity of 73 fF/με (corresponding to $8.5 \times 10^3$ ppm/με), more than 100 times larger than MEMS strain sensors reported in the literature [190]. The sensor was used to register minute fingertip and wrist dilations in response to the heartbeat at a rate of 2 kHz.

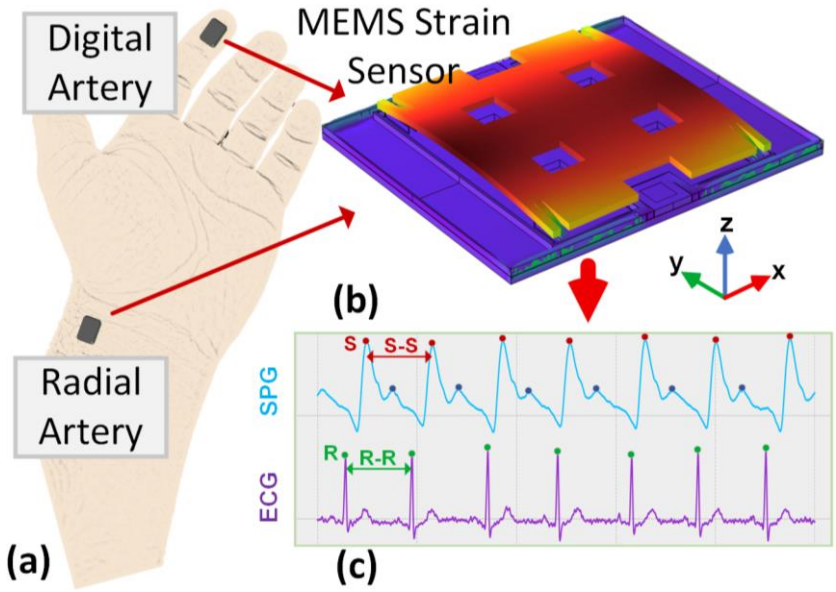


Fig. 3.11 The experimental setup of strain plethysmography (SPG). (a) The placement of the sensor on the fingertip and wrist for collecting pulse information. (b) The configuration of the sensitive strain sensor for data collection. (c) An example of the SPG signals recorded from the radial artery.

Experiments were conducted on 15 subjects, consisting of twelve males and three females. The average (± standard deviation) age, height, weight, and body mass index (BMI) of the subjects were 25.4 (± 4.04) years, 173.86 (± 7.59) cm, 76.53 (± 14.70) kg, and 25.14 (± 3.86) $kg/m^2$, respectively. The experiment began with each subject stair-stepping for 5 minutes as a strategy to elevate their heart rate. Immediately afterwards, the subject was seated on a chair in a stationary, upright position for five minutes during which the strain plethysmography signal was recorded from the fingertip at a sampling rate of 1 kHz. The subject then performed another five minutes of stair-stepping which was followed by 5 minutes of recording strain plethysmogram from the wrist. Additionally, a single-lead ECG sensor node [225] (Shimmer Sensing, Dublin, Ireland) was strapped around the chest wall with the electrodes attached to the right arm (RA), left arm (LA), and right leg (RL), to simultaneously record the RA-LA ECG signal at a sampling rate of 512 Hz as a reference for beat-to-beat intervals. The collected data were transferred to a computer for processing and analysis. The human subject experimental protocol was approved by the Institutional Review Board of Stevens Institute of Technology under protocol number 2022-044 (N).

Fig. 3.11 (c) illustrates an example of the radial SPG recording alongside its corresponding ECG signal. The beat-to-beat intervals are denoted as S-S intervals in SPG recordings and R-R intervals in ECG recordings. Within each SPG cycle, two local maxima can be observed indicating the onsets of systole (highlighted by red circles) and diastole (highlighted by blue circles). The high morphological consistency between SPG and ECG signals in capturing heartbeat information suggests that SPG can serve as a suitable alternative to ECG in monitoring HRV.

### 3.4.1.2 Signal Annotation and Parameter Extraction

To remove baseline wander and noise components, a zero-phase 2nd order Butterworth filter was used to bandpass-filter the SPG and ECG recordings within the ranges of 0.8-25 Hz and 1-50 Hz, respectively. The SPG and ECG recordings of each subject were then segmented into 120-second overlapping segments with 10 seconds of overlap between consecutive segments. ECG and SPG segments were annotated to derive R and S peaks using the Pan-Tompkins technique [90] and a custom-designed algorithm, respectively. The SPG annotation algorithm initially sets negative values to zero within

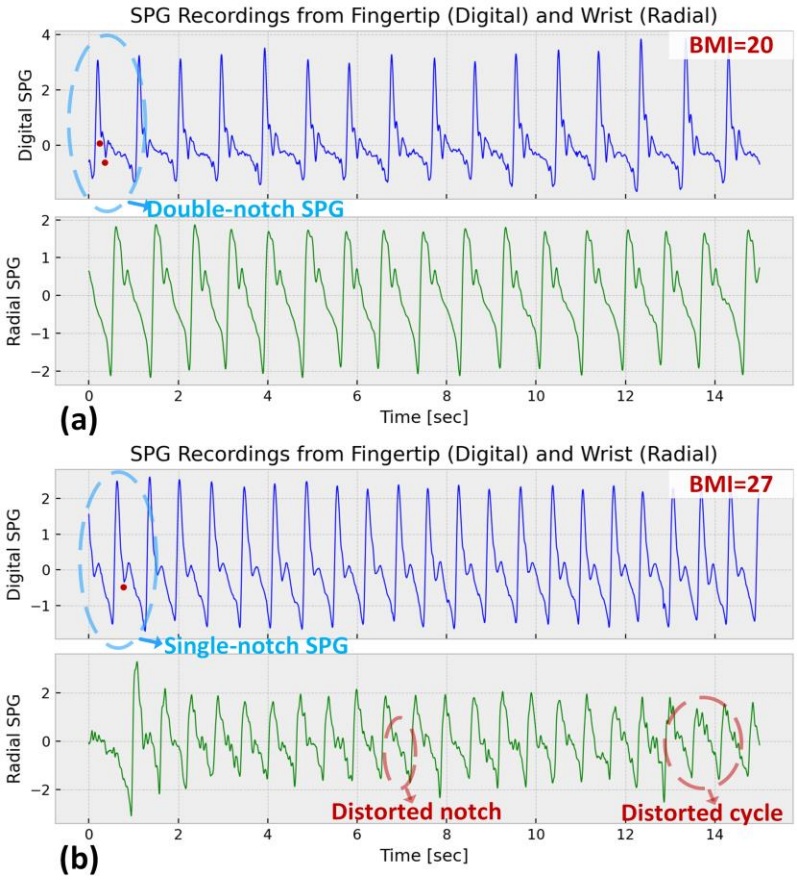


Fig. 3.12 Digital and radial SPG recordings for (a) a normal subject with BMI of 20 and (b) an overweight subject with BMI of 27.

each 120-second segment. The resulting signal is then squared to increase the prominence of the systolic peaks compared to noise and low-frequency components. The algorithm then computes two moving averages of the squared signal, one with a window length of 110 ms (110 samples) and the other with a window length of 667 ms (667 samples) resulting in $MA_1(t)$ and $MA_2(t)$ signals, respectively. These aid in determining the strain signal baseline between beats and within shorter durations. The algorithm then establishes a threshold for comparison with $MA_1(t)$:

$$\gamma(t) = MA_2(t) + 0.02 \times \overline{S^2(t)}, \qquad (3)$$

where $\overline{S^2(t)}$ denotes the average value of the squared segment. The points in $MA_1(t)$ exceeding the threshold are labeled as systolic peaks in $MA_1(t)$. Finally, the algorithm labels systolic peaks by identifying local maxima in the original strain signal. To ensure accuracy, a minimum delay of 480 ms (480 samples) is enforced between detected peaks, eliminating peaks that are too closely spaced together. The SPG and ECG segments and their respective S and R points are used to estimate HRV parameters such as mean and standard deviation of consecutive normal-to-normal (NN) heartbeats (meanNN and SDNN, respectively). Furthermore, the root mean square of successive differences (RMSSD) and the proportion of adjacent NN intervals that differ by more than 50 milliseconds (PNN50) are calculated. Along with time-domain HRVs, LF/HF: the ratio of signal energy within the low-frequency range (sympathetic activities; 0.04 - 0.15 Hz) to energy in the high-

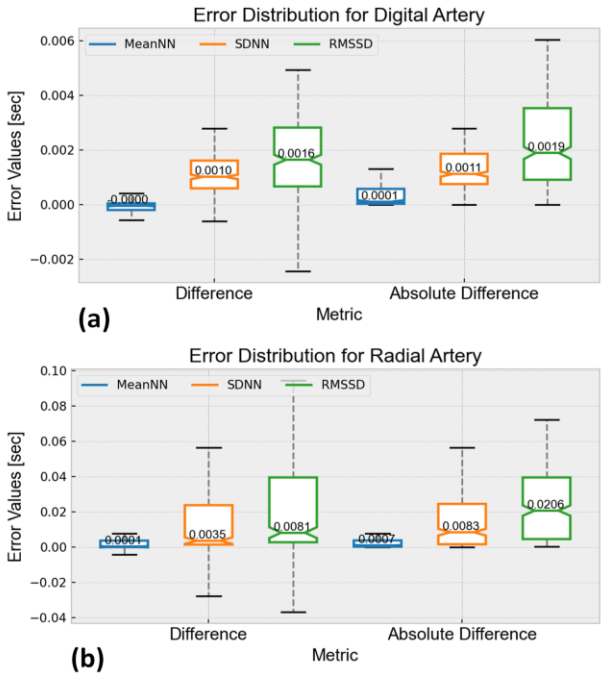


Fig. 3.13 Mean difference and mean absolute difference for meanNN, SDNN, and RMSSD values estimated from (a) digital artery and (b) radial artery.

frequency range (parasympathetic activities; 0.15 - 0.4 Hz) is calculated. This parameter indicates the balance between sympathetic and parasympathetic activities.

### 3.4.2 Experimental Results and Discussion

#### 3.4.2.1 Digital and Radial SPG Morphologies

Figs 3.12 (a) and (b) illustrate examples of digital and radial SPG recordings from subjects with BMIs of 20 and 27, respectively. Visual inspection of the figures reveal two significant findings. Firstly, in contrast to PPG recordings which have only a single notch, the SPG recordings of some of the subjects show two notches (red circles in Fig. 3.12 (a)). This phenomenon could introduce a new feature of pulse signals, and may hold clinical significance for the estimation of hemodynamic parameters such as blood pressure. Second, it was observed that as the BMI increases, the cycle morphology and signal

notches in the radial artery (wrist) may distort. This observation can be explained by the higher concentration of fat tissues in the wrist.

### 3.4.2.2 Performance for SPG-based HRV Estimation

Statistical analyses based on mean difference and mean absolute difference were conducted to evaluate the performance of HRV estimated from SPG and corresponding ECG recordings. Figs 3.13 (a) and (b) illustrate the error distributions for digital and radial arteries. The median differences (absolute differences) of 0.0 (0.7) ms, 1 (1.1) ms, and 1.6 (1.9) ms were reported for the estimation of meanNN, SDNN, and RMSSD, from the digital artery respectively. These error values were reported for the radial measurements as 0.1 (0.7) ms, 3.5 (8.3) ms, and 8.1 (20) ms for meanNN, SDNN, and RMSSD, respectively.

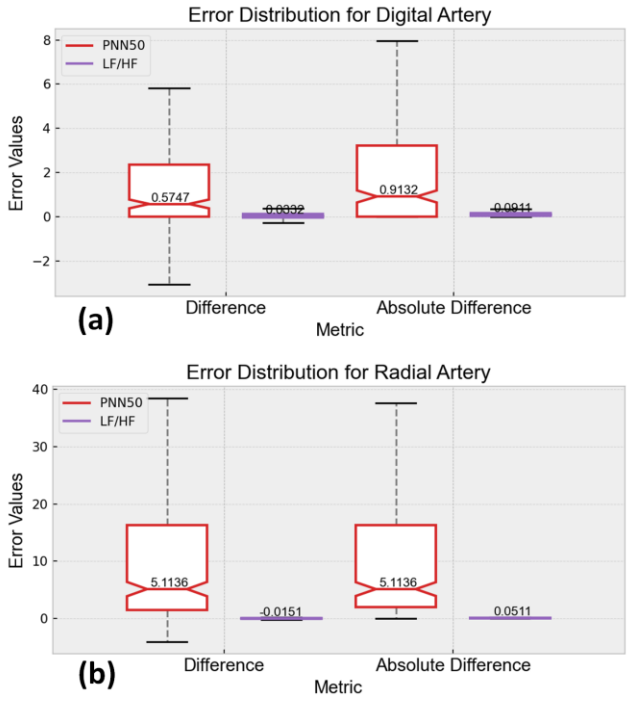


Fig. 3.14 Mean error and mean absolute error for PNN50 and LF/HF values estimated from (a) digital artery and (b) radial artery.

The comparison between the digital-derived and radial-derived HRVs suggests that the fingertip is a more reliable site for continuous monitoring of HRV parameters.

Figs 3.14 (a) and (b) depict the difference and mean difference of PNN50 and LF/HF derived from SPG and ECG for digital and radial arteries, respectively. Comparing the differences achieved for PNN50 of digital and radial arteries (0.57% vs. 5.11%), it is demonstrated that the cycle distortion observed in Fig. 3.12 (b) has hindered the algorithm to identify the systolic peaks (S) correctly. The median difference of LF/HF, on the other hand, shows excellent consistency between ECG and SPG recordings for both digital and radial arteries (0.0332 and -0.0151). This implies that SPG-based pulse information could

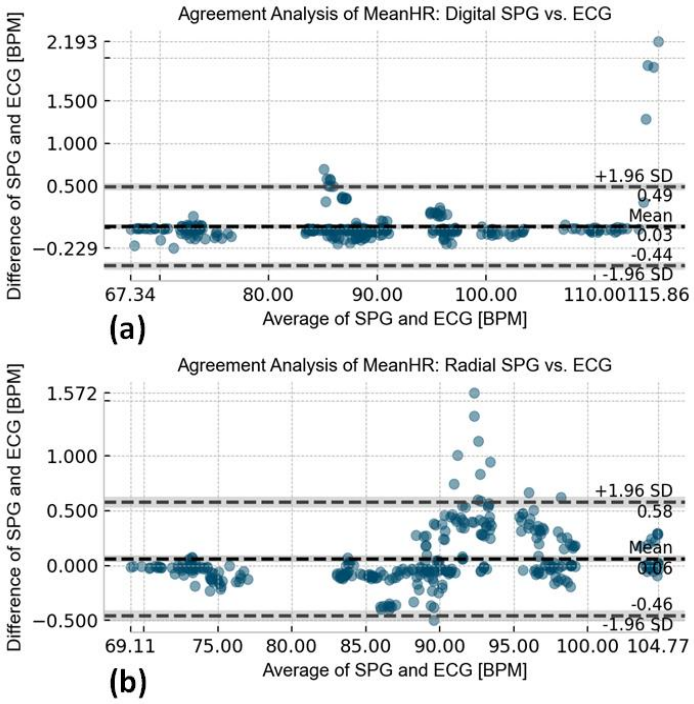


Fig. 3.15 Bland-Altman agreement between meanHR estimated from (a) digital and (b) radial arteries, with respective ECG-derived meanHR values.

be a reliable means for assessing the balance between sympathetic and parasympathetic activities.

The statistical agreement between the estimated and reference meanHR values for digital and radial arteries are investigated in Bland-Altman characteristics with confidence intervals of 95% (±1.96 × SD) in Figs 3.15 (a) and (b), respectively. According to these figures, the majority of differences for digital and radial-derived meanHR fall within the agreement limit ranges of [-0.44, 0.49] BPM and [-0.46, 0.58] BPM, respectively. Furthermore, the mean differences of 0.03 BPM and 0.06 BPM suggest negligible estimation biases for average HR in 120-second segments, with digital recordings offering superior performance for HRV monitoring. This finding is consistent with median differences achieved in Figs 3.13 and 3.14.

#### 3.4.2.3 Comparative Analysis

The proposed SPG technology is compared with monitoring systems based on chest-worn seismo-cardiogram (SCG) [232] and wrist-worn photo-plethysmogram (PPG) [233], for meanNN estimation in Table 3.4. As can be seen, the analyses were performed based on the regression characteristic in which the predictability of ECG-derived meanHR from SPG-derived meanHR is assessed. The least root mean square errors (RMSE) belong to the wrist PPG with 4.01 ms and digital SPG with 4.2 ms, whereas the worst performance was reported for the SCG-based method with RMSE of 13 ms, suggesting a superior performance for the SPG-based method. Furthermore, the coefficient of determination ($r^2$) metric ranks the digital SPG as the top-performing HRV estimation method with an $r^2$ of 0.998 which is consistent with the intercept of the fitted lines for digital SPG and chest

Table 3.4 Comparative Study with Chest SCG and Wrist PPG

| Method | Regression characteristic | | | Bland-Altman | |
|---|---|---|---|---|---|
| | RMSE (ms) | $r^2$ | $y = ax + b$ | MD (ms) | LOA [U, L] (ms) |
| **Digital SPG** | **4.2** | **0.998** | **1.004$x$-0.004** | **-0.00** | **20 [-10, 10]** |
| Radial SPG | 12.5 | 0.977 | 1.051$x$-0.039 | -0.00 | 50 [-30, 20] |
| Chest SCG [232] | 13 | 0.99 | 1.01$x$-6.29 | 0.37 | 52 [-26, 26] |
| Wrist PPG [233] | 4.01 | - | - | 1.17 | - |

SCG (-0.004 vs -6.29, respectively). The bias of 0.0 ms for both digital and radial-derived meanHR denotes superior performance compared to SCG and PPG methods with estimation biases of 0.37 ms and 1.17 ms respectively. The smallest limit of agreement (LOA) belongs to digital SPG-based meanHR, implying its superior performance compared to radial SPG and chest SCG with LOAs of 50 ms and 52 ms, respectively.

## 3.5 Prospects of Cuffless Pulse Pressure Estimation from a Chest-Worn Accelerometer Contact Microphone

Cardiovascular diseases (CVDs) are the leading cause of mortality in the world, accounting for 32% of deaths in 2019 [234]. Early detection of CVD symptoms allows for timely risk management with proper interventions. Pulse pressure (PP), defined as the difference between the systolic and diastolic blood pressures, is among the most significant indicators of CVDs such as coronary heart disease [235].

Current blood pressure (BP) monitors employ an inflatable cuff to measure systolic and diastolic blood pressures. However, the intrusive nature of cuff-based BP monitors inhibits their use as a ubiquitous monitoring device. Recent advances in wearable sensors over the past decade have revolutionized blood pressure monitoring [236]. Several research works have addressed blood pressure estimation by introducing surrogates of pulse transit

time (PTT) [237] or pulse arrival time (PAT) [238] using an electrocardiography (ECG) or seismo-cardiography (SCG) sensor in conjunction with a photoplethysmography (PPG) sensor attached to the earlobe, fingertip, or toe. The simultaneous use of two sensors however, causes inconvenience for the subject, hence provoking the need for less-intrusive devices. On the other hand, the precision accelerometer contact microphone introduced recently in [208] has shown promise in identifying cardiovascular abnormalities such as peripheral artery disease [132], ejection fraction [222], and valvular heart diseases [220,221], using a single sensor setup for recording thoracic recordings.

As demonstrated in [239], pulse pressure is proportional to the stroke volume, which is defined as the blood volume pumped from the heart per cycle. In this study, we assume that changes in the stroke volume appear as changes in the energy level of heartbeat-induced vibrations on the chest wall. Hence, measuring these changes could potentially enable the estimation of pulse pressure. In this work, we estimate pulse pressure using a single sensitive accelerometer contact microphone (ACM) placed on the chest wall.

### 3.5.1 Experimental Setup and Methodology

#### 3.5.1.1 Data Collection

In this work, twenty subjects with cardiovascular diseases (CVDs), consisting of nine male and eleven females, are studied at the cardiac care unit of Sorin Medical P.C. in New York City. The average (± standard deviation) age, height, and weight of the patients are 70.25 (± 10.77) years, 167.04 (± 12.09) cm, and 82.32 (± 22.07) kg, respectively.

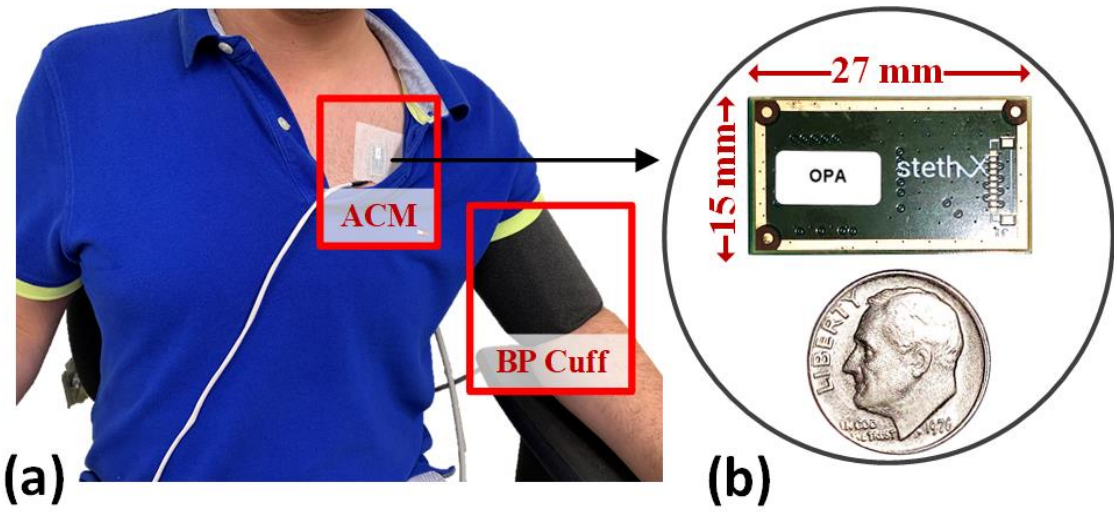


Fig. 3.16 (a) Experimental setup including an accelerometer contact microphone (ACM) attached to the pulmonary region and an expanded-coverage, preformed blood pressure cuff (OMRON HEM-FL31) secured around the left arm as reference. (b) The dimensions of the ACM sensor head.

As shown in Fig. 3.16 (a), the experimental setup consists of a $\pm 4$ g sensitive accelerometer contact microphone (ACM) with micro-g resolution and small form-factor of 27 mm × 15 mm × 2.5 mm (Fig. 3.16 (b)) from StethX Microsystems [128], which is attached to the pulmonary region of the subjects using medical-grade adhesive tape. This device is a low-noise accelerometer with a wide operational bandwidth of 0-10 kHz, allowing for recording heartbeat-induced sounds and vibrations on the chest wall. The device is not sensitive to airborne emission sounds, making it a robust phonocardiogram sensor against acoustic ambient noise. The subjects were seated at rest on a chair for a period of five minutes, followed by five minutes of ACM measurements at a sampling rate of 22.33 kHz, still at a seated position. Immediately afterwards, an OMRON-BP786N blood pressure monitor was used to measure the systolic blood pressure (SBP) and diastolic blood pressure (DBP) of the subjects from their left arms. Pulse pressure was then

calculated by subtracting the DBP from SBP. The patient experimental protocol was approved by the Institutional Review Board of Stevens Institute of Technology under protocol number 2022-044 (N). The collected data were transferred to a computer and processed in a Python framework.

### 3.5.1.2 Signal Pre-processing

ACM recordings were initially band-pass filtered using a zero-phase, 3rd-order Butterworth filter over the frequency range of 2 Hz - 4 kHz to remove respiration artifacts and baseline wander. Filtered signals were then down-sampled to a sampling rate of 4 kHz to reduce the computational complexity of the processing steps. Subsequently, ACM signals were segmented into 10-second time frames similar to our previous work [125]. The data segmentation procedure was carried out twice, employing two different approaches. The first approach involved 90% overlap between consecutive segments,

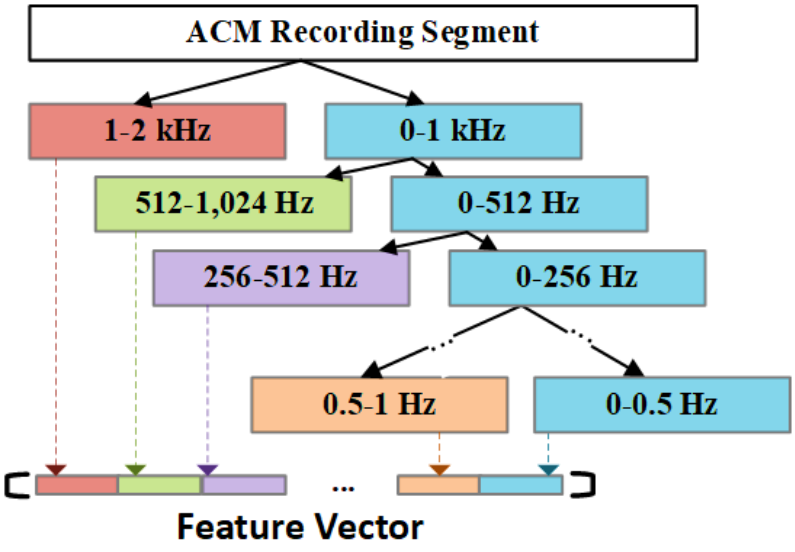


Fig. 3.17 Feature extraction using signal decomposition based on discrete wavelet transform (DWT). Wavelet sub-bands are used for feature extraction.

while the second approach utilized a 10-second gap between segments. The latter strategy was performed to prevent data leakage and ensure the integrity of the results. Finally, segments contaminated with motion artifacts due to movements during recordings were removed from ACM signals by applying a root-mean-square (RMS) filter with a sliding window of 500 ms for signal segmentation. The segment removal threshold was selected as twice the median value of the filter, setting the criterion for discarding the segments which exceeded the threshold level. The resulting motion-free signal segments were used for feature extraction.

#### 3.5.1.3 Feature Extraction from ACM Recordings

As mentioned earlier, variations in heartbeat-induced vibrations on the chest wall contribute to pulse pressure estimation. The vibration on the chest wall can be directly recorded by the ACM, enabling the investigation of time-domain features. In order to characterize the ACM recordings, discrete wavelet transform (DWT) with a symlet kernel is employed to decompose the segments into 12 sub-bands (levels) for feature extraction, as illustrated in Fig. 3.17. At each decomposition level, the segments are split into high-frequency and low-frequency time series with coefficients called detail coefficients (DCs) and approximation coefficients (ACs) respectively, as depicted in Fig. 3.18.

In this work, we extract features from 12 detail coefficient sets corresponding to 0.5-1 Hz, 1-2 Hz, 2-4 Hz, 4-8 Hz, 8-16 Hz, 16-32 Hz, 32-64 Hz, 64-128 Hz, 128-256 Hz, 256-512 Hz, 512-1024 Hz, and 1-2 kHz. Additionally, features from the coefficient set associated with 0-0.5 Hz are appended to the feature vector to provide a comprehensive representation of the segment within the whole frequency range. Per signal segment,

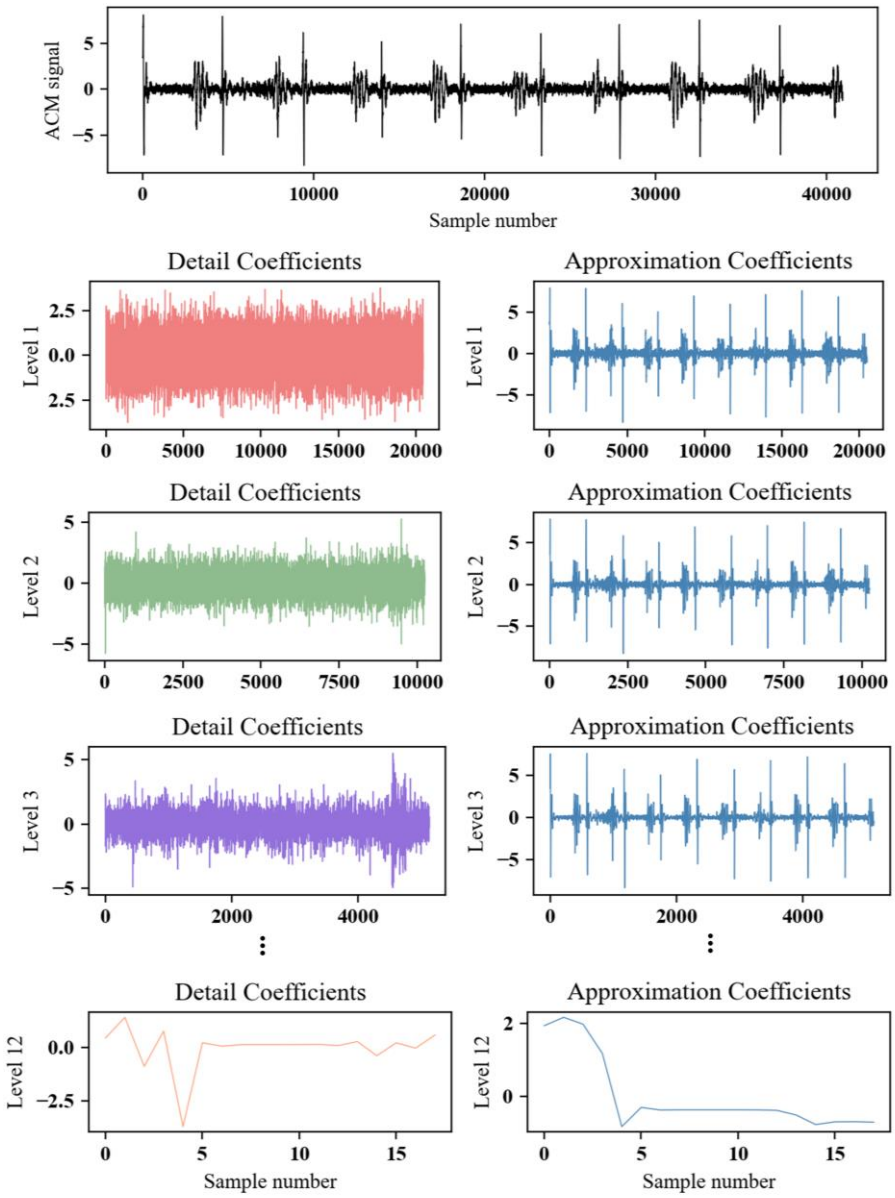


Fig. 3.18 Detail coefficienis (DCs) and approximation coefficients (ACs) of a 10-sceond ACM signal segment achieved by 12-level DWT decomposition.

features include the mean, standard deviation, skewness, kurtosis, 10th percentile, 25th percentile, 50th percentile, 75th percentile, and 90th percentile of the absolute and squared values of 13 coefficient sets (i.e., $13\times 2\times 9$ features). Furthermore, the differences

between every pair of the afore-mentioned percentiles are incorporated in the feature space (i.e., $13\times 2 \times 10$ features). As such, each 10-second signal segment is represented by a total of 494 features.

#### 3.5.1.4 Regression Model for Pulse Pressure Estimation

According to our sample size, a machine learning regression model based on the extreme gradient boosting (XGBoost) method was implemented to estimate pulse pressure values from the extracted features. This model employs ensemble decision trees to minimize the regression error. XGBoost was trained on the sample space to minimize the square error loss function between the predicted and ground-truth pulse pressures. The model was optimized using the gradient descent method.

### 3.5.2 Experimental Results and Discussion

A total of 474 and 8,275 segments are obtained from 20 patients with an average (± standard deviation) pulse pressure of 49.15 (± 11.83) mmHg for non-overlapped and 90%-overlapped data, respectively. It has been demonstrated that pulse pressure is inversely proportional to the pulse arrival time (PAT) [239]. In this study, since the subjects

Table 3.5 Performance Evaluation of the Pulse Pressure Estimation Framework.

| **Metric** | **Blood Pulse Pressure Estimation Error (± Standard Deviation)** | |
|---|---|---|
| | **Leave-Subject-Out Validation** | **Cross-Validation (10-Fold)** |
| **RMSE (mmHg)** | 11.41 (± 6.42) | 2.15 (± 0.71) |
| **MAE (mmHg)** | 10.49 (± 6.73) | 1.87 (± 0.38) |
| **Acc. (%)** | 77.14 (± 19.09) | 95.68 (± 1.17) |

are at rest, PAT is assumed to be constant over the short duration of measurement. As such, the segments corresponding to each subject are labeled by the same pulse pressure value calculated as the difference between the measured SBP and DBP. Two validation strategies are adopted to evaluate the performance of the predictive model. The first strategy is 20-fold leave-subject-out validation (LSOV), where the model is trained on the data of 19 subjects and tested on the remaining held-out subject. The second strategy is 10-fold cross-validation (10-CV), where the non-overlapped data is divided into two parts; 426 segments (90%) are used for training and the remaining 48 segments (10%) constitute the test set. The LSOV method differs from 10-CV in the sense that LSOV considers the dependence among the segments corresponding to the same subject, whereas 10-CV assumes that segments of the same subject are independent. Although 10-CV has been widely used in the literature, LSOV offers a more realistic evaluation.

#### 3.5.2.1 Pulse Pressure Estimation Evaluation

Table 3.5 summarizes the performance of the proposed pulse pressure estimation framework in terms of root mean square error (RMSE), mean absolute error (MAE), and accuracy (Acc.), defined as the percentage of relative error. Average RMSE values of 11.41 mmHg and 2.15 mmHg for LSOV and 10-CV, respectively, suggest robust estimation of pulse pressure. Similarly, MAE values of 10.49 mmHg and 1.87 mmHg respectively reported for LSOV and 10-CV demonstrate high linear consistency between the actual pulse pressure and associated predictions. This consistency is further confirmed by average accuracies of 77.14% and 95.68% achieved for LSOV and 10-CV, respectively. The

Table 3.6 Contributive Frequency Bands for Pulse Pressure Estimation.

| **Number** | **Sub-bands Contributions** | |
|---|---|---|
| | **Band Score** | **Frequency Sub-band** |
| **1** | **156** | **1-2 kHz** |
| **2** | **37** | **512–1,024 Hz** |
| **3** | **19** | **64–128 Hz** |
| **4** | **17** | **128–256 Hz** |
| **5** | **14** | **32–64 Hz** |
| **6** | **11** | **16–32 Hz** |
| 7 | 7 | 256–512 Hz |
| 8 | 7 | 4–8 Hz |
| 9 | 6 | 2–4 Hz |
| 10 | 5 | 8–16 Hz |
| 11 | 2 | 0.5–1 Hz |
| 12 | 1 | 1–2 Hz |
| 13 | 1 | 0–0.5 Hz |
| **Average** | **21.76** | **N/A** |

relatively small difference between the performance achieved for LSOV and 10-CV suggests that the predictive model could take advantage of pre-knowledge about the subject's cardiac cycle and the corresponding pulse pressure values for the accurate estimation of pulse pressure. This is potentially obtained by fine-tuning the predictive model on a few seconds of the subject's recordings and the corresponding pulse pressure readings.

The authors in [240] have reported average RMSEs of 3.59 mmHg and 2.24 mmHg for the static squat exercise and the cold pressor test, respectively, based on 5-fold cross-validation of 11 healthy subjects. Considering the 2.15 mmHg RMSE achieved on 20 CVD patients in this study, the proposed pulse pressure estimation framework shows promise for CVD patients with wide ranges of PP. The results of this study were published in [241].

#### 3.5.2.2 Significant Frequency Sub-Bands

The contribution of each frequency sub-band to pulse pressure estimation is reported in Table 3.6. For each subject in LSOV, the features are ranked based on their occurrence frequency in the ensemble trees. A feature is selected as highly contributive if its frequency exceeds the following threshold:

$$\tau = \sigma + \delta, \qquad (4)$$

where $\sigma$ and $\delta$ represent the average and standard deviation of the frequency distribution, respectively. The top features from all subjects are then combined, and the frequency of each feature is calculated again. Finally, a band score is calculated based on the occurrence of each frequency sub-band in the combined list as summarized in Table 3.6. According to this table, the most and least contributive frequency bands are 1-2 kHz and 0-0.5 Hz with scores of 156 and 1, respectively. It is also demonstrated that the medium frequency band 64-128 Hz with a band score of 19 suggests higher correlation with PP compared to high-frequency bands such as 256-512 Hz with a band score of 7.

# 4 Fetal Heart Rate Estimation Using Vibrational and Biopotential Sensors

## 4.1 Fusion of Biopotential and Vibrational Sensors for Fetal Heart Rate Estimation

Ultrasound cardiotocography (CTG) enables the continuous monitoring of fetal heart sound through auscultation, and provides obstetricians with information regarding FHR and uterine contractions [242]. CTG requires pregnant individuals to visit the clinic intermittently. Despite being a non-invasive modality, CTG has not conclusively been proven to be a safe technology as it irradiates the fetus with ultrasound frequencies [243]. Furthermore, it is prone to missing information due to fetal movement during monitoring [46]. Other than CTG, FHR monitoring is carried out through fetal electrocardiography (fECG) by placing low-noise ECG electrodes on the abdomen of the mother [46]. Consequently, a mixture of maternal ECG (mECG) and fECG is achieved which is named abdominal ECG (aECG). Over the past decade, a huge amount of research, primarily based on blind source separation (BSS) [244,245] and adaptive filtering [47,246], has been dedicated to extracting fECG imposed by noise components with signal-to-noise ratio (SNR), occasionally as low as -20 dB [247].

Recently, wearable seismo-cardiography (SCG) and gyro-cardiography (GCG) have been widely employed for wearable heart monitoring and disease detection. Our research group has reported on the application of SCG and GCG modalities on FHR estimation [43], where three inertial sensors were used to measure the abdominal movements caused by fetal heartbeats. Although the potentiality of FHR extraction was demonstrated, the reported PPA was 75.20% since the vibrations caused by fetal cardiac activity may not fully transfer to the abdomen, especially from the 28th to the 37th week of gestation due to

the greasy layer called vernix, which dampens the vibrations produced by the heartbeat of fetus [60].

In this work, a multi-modal framework is developed where ECG and SCG/GCG modalities are employed in a fusion-based context for FHR extraction. It is assumed that data fusion among these modalities would enhance the appearance of fetal heartbeat components. The proposed framework employs off-the-shelf sensors, not necessarily designed for low-noise measurements. To the best of our knowledge, this is the first study addressing the fusion between ECG and SCG/GCG modalities for fetal heart rate estimation.

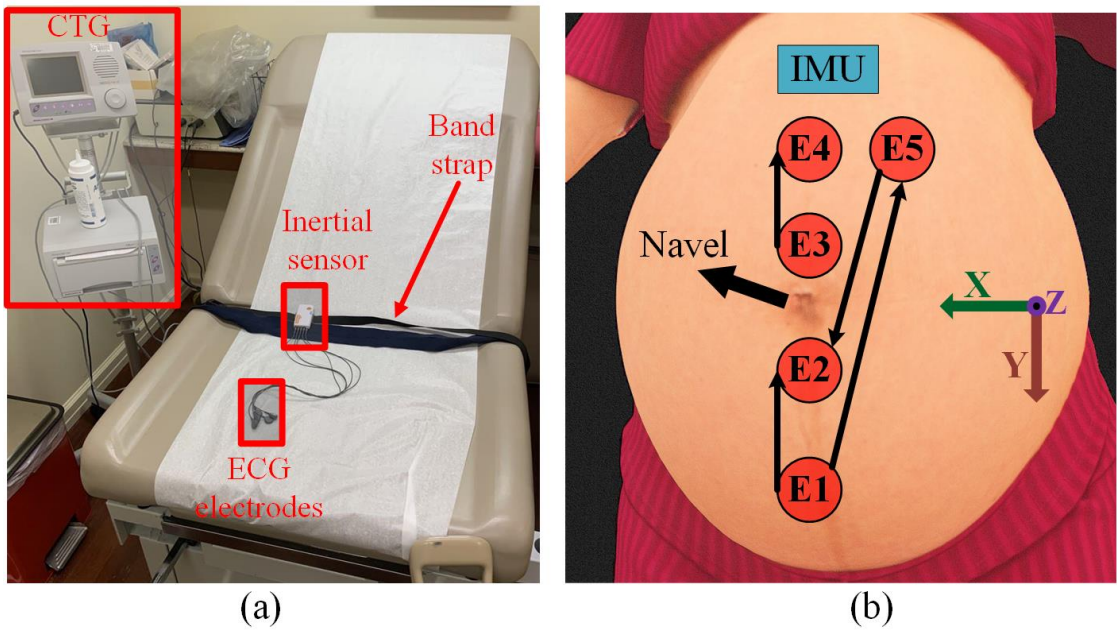


Fig. 4.1 (a) The experimental setup for data collection: an adjustable bed, the wearable sensors, CTG machine, and band straps. (b) ECG electrodes and IMU layout on the pregnant abdomen; X, Y, and Z show the axes for IMU recordings.

### 4.1.1 Methodology

In the following sub-sections, the experimental setup, signal processing, and sensor fusion techniques are explained in detail.

#### 4.1.1.1 Experimental Setup and Measurement Protocols

Ten normal healthy pregnant women participated in this study and fetal heart rates were measured at the Department of Obstetrics and Gynecology at New York University Grossman School of Medicine (NYUGSM) after obtaining informed consent. The patients' experimental protocol was approved by the Institutional Review Board (IRB) of NYUGSM under study number i18-00564. The subjects' average (standard deviation) gestational age was 37.11 (2.56) weeks, with body mass index (BMI) falling within the range of $28.91 \pm 4.46$ Kg/m$^2$. Also, the average (standard deviation) age of the subjects was 32.8 (4.27) years old. Fig. 4.1 depicts the experimental setup.

The subjects were asked to lie down on an examination table, shown in Fig. 4.1 (a), for 5 minutes in a supine position. Then, an inertial measurement unit (IMU) from Shimmer Sensing was secured around the top part of the subject's abdomen using a band strap. The IMU sensor was used to record the rotational and linear vibrations of the abdomen. Furthermore, the electrodes of a four-lead ECG sensor were placed along the navel-to-chest direction as demonstrated in Fig. 4.1 (b). The layout of ECG electrodes (E1-E5) was designed such that they covered as many paths (purple arrows) on the abdomen as possible. According to Fig. 4.1 (b), the X, Y, and Z axes correspond to the directions along right-to-left, top-to-bottom, and dorso-ventral, respectively. All sensors recorded the data at a sampling rate of 512 Hz. In addition, an ultrasound CTG machine was employed to record

the ground-truth FHR. After the measurement, the recorded data was transferred to a computer for further processing.

#### 4.1.1.2 Signal Pre-Processing and EMD Refinement

In order to remove the baseline wandering and motion artifacts from the recordings, ECG and SCG/GCG signals were filtered using a third-order zero-phase Butterworth band-pass filter with cut-off frequencies of 0.8-40 Hz and 0.8-25 Hz, respectively. The achieved signals are assumed to contain components associated with fECG, mECG, and noise, which should be separated from one another. On the other hand, mechanical activities, i.e., SCG/GCG, are delayed compared to electrical activities, i.e., ECG, in the heart, provoking the need for spectral analyses rather than time-domain methods. To this end, we designed an algorithm based on empirical mode decomposition (EMD), coined EMD refinement (EMDR), to keep those components that are highly correlated with fetal cardiac activities. EMDR starts with decomposing a signal into its intrinsic mode functions (IMFs), the top ones of which, carrying important information of the cardiac activity, are kept, whereas the remaining IMFs are discarded. In this work, the first six IMFs are kept as they comprise higher frequency components. Each $\mathrm{IMF_i}$ is scored by a factor of $\xi_i (i = 1,2,3,4, \dots)$, which is defined as follows:

$$\xi_i = \left( \sum_{k=\lfloor \frac{2N}{fs} \rfloor}^{\lfloor \frac{3N}{fs} \rfloor} |A_i[k]|^2 \right) / \left( \sum_{k=1}^{N/2} |A_i[k]|^2 \right), \quad (1)$$

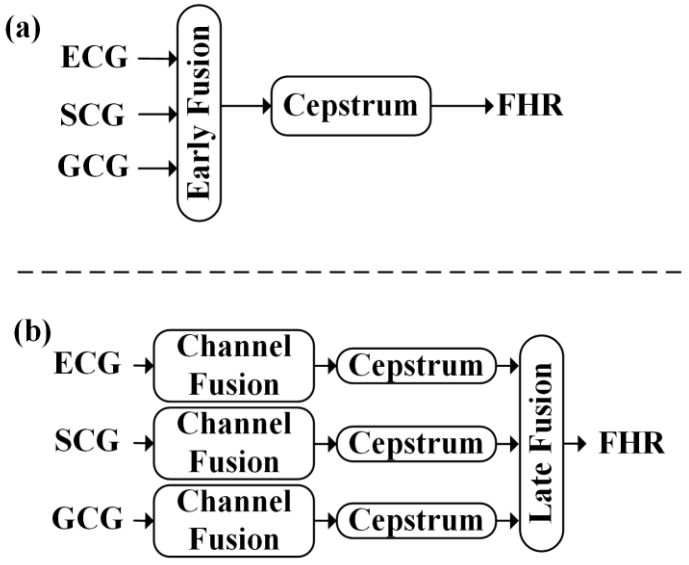


Fig. 4.2 Fusion-based FHR extraction flow graphs: (a) Early fusion, (b) Late fusion.

where $A_i[k]$ denotes the representation of i-th IMF in the spectral domain, $N$ shows the signal length, and $f_s$ implies the sampling frequency of the signal. The score for each IMF is determined based on the proportion of the information concentrated in a specific range to the total energy of the IMF. According to the Nyquist rate and assuming that the sample number $k = \frac{N}{2}$ in FFT domain corresponds to $f = \frac{f_s}{2}\ Hz$ in the frequency domain, $2N/f_s$ and $3N/f_s$ would imply 2 and 3 Hz in the frequency scale, respectively. This means that the IMF's are scored according to a frequency range, at which most of the energy of fetal FHR components, 120-160 BPM, exists. Thus, by scoring the IMF's, they can be ranked based on their information of fetal cardiac activity. Hence, considering the top scores as $\hat{\xi}_j$ ($\hat{\xi}_j > \hat{\xi}_{j+1}$) and their respective IMF's as $\widehat{IMF}_j$, the refined signal $\hat{x}$ would be reconstructed by using the top four IMF's out of six as follows:

$$\hat{x} = \hat{\xi}_1 \times \widehat{IMF}_1 + \hat{\xi}_2 \times \widehat{IMF}_2 + \hat{\xi}_3 \times \widehat{IMF}_3 + \hat{\xi}_4 \times \widehat{IMF}_4. \quad (2)$$

Hence, EMDR could improve the signal components at a certain frequency range. EMDR is applied on every channel of ECG, SCG, and GCG, turning the signals more suitable for the multi-modal signal fusions illustrated in Fig. 4.2. As shown in this figure, two fusion techniques are designed: early-fusion (Fig. 4.2 (a)) and late-fusion (Fig. 4.2 (b)) as elucidated in the following sections.

#### 4.1.1.3 Early-Fusion FHR Estimation

In order to implement a meaningful fusion among the modalities, we employ the spectral contents of the signal since the nature of ECG, SCG, and GCG are different from one another in the time domain. Furthermore, having considered the signal refinement proposed above, the enhanced frequency components within the range of 2-3 Hz are deemed to contribute to a constructive effect in the spectral domain. The time-frequency (TF) representation of each channel is calculated by applying a Morse continuous wavelet transform (CWT) as defined below:

$$\Upsilon_{P,\upsilon}(\omega) = U(\omega) a_{P,\upsilon} \omega^{\frac{P^2}{\upsilon}} \times \exp(-\omega^{\upsilon}), \quad (3)$$

where $U(\omega)$, $P^2$, $\upsilon$, and $a_{P,\upsilon}$ characterize the unit step function, time-bandwidth product, symmetry of the Morse wavelet, and normalizing constants, respectively. In this study, the values of $\upsilon$ and $P$ are set to 4 and 120, respectively. Once the refined signals from each modality are transformed into their TF representation, the values of each TF are normalized to a zero-mean unit-variance distribution, i.e., $\sim\mathcal{N}(0,1)$. Then, the normalized TF's are averaged within their respective modalities, normalized again, and averaged across the

three modalities. Indeed, the fusion operator generates a single TF characteristic that represents the combined information from the three modalities with an emphasis on the components within 2-3 Hz, which is relevant to fetal heart activities. An inverse continuous wavelet transform (ICWT) is then performed on the fused TF to convert it back to time domain. Subsequently, to obtain the FHR values, cepstrum is performed on the signal to calculate the lag at which the time-domain signal appears to have a higher energy. As the FHR value varies within the range of 110-180 BPM throughout the entire experiment, the desired peak is to be detected within the corresponding lag values, i.e., 333-545 ms. Hence, FHR is calculated as follows:

$$\text{FHR (BPM)} = \frac{60(second)}{lag\ (second)}. \quad (4)$$

Table 4.1 RMSE of Estimation by Modality-only and Fusion-based Methods

| Subject | SCG | GCG | ECG | Early fusion | Late fusion |
|---|---|---|---|---|---|
| **1** | 12.07 | 11.69 | 12.65 | 12.37 | **9.52** |
| **2** | 16.25 | 17.15 | 15.85 | 16.04 | **14.70** |
| **3** | 13.10 | 12.01 | 14.22 | 12.01 | **11.30** |
| **4** | 12.07 | 11.86 | 11.44 | 12.81 | **9.22** |
| **5** | 17.31 | **12.34** | 13.64 | 14.16 | 12.54 |
| **6** | 8.34 | 11.26 | 10.11 | 9.77 | **7.34** |
| **7** | 14.62 | 14.20 | 14.91 | 16.43 | **12.65** |
| **8** | 12.75 | 12.38 | 12.69 | 12.14 | **10.00** |
| **9** | 12.74 | 12.54 | 9.85 | **9.42** | 9.68 |
| **10** | 10.54 | 10.56 | 10.59 | 9.52 | **8.57** |
| **Mean (± std.) (BPM)** | 12.98 (± 2.47) | 12.60 (± 1.76) | 12.60 (± 1.96) | 12.47 (± 2.38) | **10.55 (± 2.10)** |

#### 4.1.1.4 Late-Fusion FHR Estimation

The late-fusion flow graph is shown in Fig. 4.2 (b). According to this structure, each modality undergoes a channel fusion procedure as described above. As such, ECG, SCG, and GCG channels are fused within their respective modalities separately. Once the fusion at each modality is performed, three TF representations corresponding to ECG, SCG, and GCG are obtained. Unlike the early-fusion technique, the TF associated with each modality is converted back to the time domain with no data fusion at this stage. Then, cepstrum is applied on each time-domain signal, and FHR values are estimated from each modality separately. Thus, three estimates of FHR are acquired through SCG, GCG, and ECG channels. Data fusion is then conducted on the three FHR estimates by averaging the values. In fact, the three modalities are assumed to introduce some amount of error in the FHR estimate values. Yet, late-fusion is meant to reduce the total error by moderating the FHR discordance among the sensors. This will be further discussed in the experimental results.

### 4.1.2 Experimental Results and Discussion

In this section, the experimental results are presented and a comparison between the proposed techniques and those in literature is made.

#### 4.1.2.1 Performance Evaluation

In order to evaluate the performance of the proposed methods, the signals are segmented into 10-second frames, from each of which FHR values are estimated. The estimate values are then compared with the ground-truth values, i.e., CTG readings, through three metrics called positive percent average (PPA), i.e., the percentage of time the

proposed method generated a valid FHR within 10% of the readings from CTG, absolute percentage error (APE), and root-mean-square error (RMSE). Table. 4.1 summarizes the RMSE results for fusion-based and single-modality scenarios for the ten subjects, where single-modalities include signal refinement and channel fusion. According to Table. 4.1, late-fusion suggests the best performance among all by 10.55 (± 2.10) BPM of RMSE, whereas early-fusion is reported by 12.47 (± 2.38) BPM, indicating superior accuracy to the single-modality scenarios such as SCG, GCG, and ECG with 12.98 (± 2.47), 12.60 (± 1.76), and 12.60 (± 1.96) BPM, respectively, where GCG introduces the least error. The difference in performance between early and late fusions results from the potentially destructive effect caused by the complex values of TF while fusing the sensors in early-fusion. Furthermore, for subjects 5 and 9, the best RMSE's were achieved by the GCG modality and early-fusion, respectively, although late-fusion performs more robustly in other cases. Fig. 4.3 makes a comparison of the APE and PPA metrics. As illustrated in

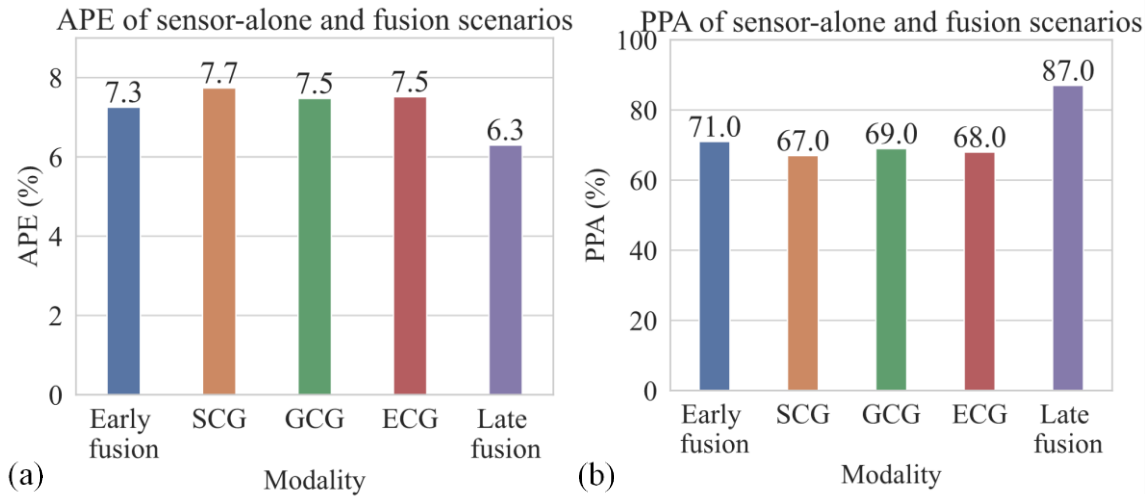


Fig. 4.3 Performance of the proposed method and the sensor-alone scenarios: (a) APE, (b) PPA.

Table 4.2 Comparison with Other Works in Literature

| Modality | PPA (%) | RMSE (BPM) | Confidence Interval | Reference |
|---|---|---|---|---|
| **Low SNR ECG/SCG/GCG** | **87.00 (± 4.81)** | **10.55 (± 2.10)** | **(-1.9, -0.19)** | **Proposed method** |
| SCG/GCG | 75.20 (± 11.81) | 11.40 (± 2.17) | N/A | [43] |
| fECG | 83.40 (± 15.40) | 4.80 (± 2.00) | N/A | [248] |
| PCG/fECG | N/A | N/A | (-8.84, 8.24) | [249] |

Figs 4.3 (a) and (b), the late-fusion model outperforms early-fusion as well as single-modality settings by a margin of 1.00% and 17.00% in terms of APE and PPA, respectively. Unlike the RMSE values mentioned in Table. 4.1, PPA=71.00% and APE=7.30% of early-fusion suggest better performances compared to the single-modality scenarios. The worst performance belongs to the SCG modality with 12.98 BPM, 7.70%, and 67.00% of RMSE, APE, and PPA, respectively. In summary, late-fusion offers a more robust and reliable estimation of FHR values.

#### 4.1.2.2 Comparison with Other Methods

As the proposed framework leverages both abdominal ECG and SCG/GCG modalities, where CTG provides the ground-truth value, it is compared with the three most similar wearable-based works in the literature. In [43], SCG/GCG are employed to extract the FHR, where CTG is used as the reference. In [248], FHR extraction is performed using a small AN24 sensor from Monica Healthcare, which was used to record fECG. Furthermore, a sensor fusion of phono-cardiography (PCG) and fECG addresses FHR extraction in [249], with which the proposed method is compared. These works were

selected for comparison since they leverage a partially-similar setup to ours, i.e., inertial sensors, abdominal ECG sensors, and a sensor fusion algorithm, respectively. Table. 4.2 reports on the performance comparison between the proposed method and other works. As seen in Table. 4.2, the proposed late-fusion framework outperforms the other methods in the literature in terms of PPA by a large margin of 4.00%, whereas in terms of RMSE, it suggests weaker performance than the fECG-based method (10.55 vs. 4.80 BPM). This weaker performance stems from using commercial off-the-shelf ECG sensors in our setting. Furthermore, despite employing a single-IMU setting, the proposed method could estimate the FHR values more accurately than the triple-IMU-based framework in [43] (RMSE: 10.55 vs. 11.40 and PPA: 87.00% vs. 75.20%). The fourth column of Table. 4.2 represents 95% confidence interval (CI) (lower bound, higher bound) for the difference between the reference and the estimates of FHR. This metric is only reported for the fusion algorithm in [249]. Comparing the proposed method and [249], it is concluded that most of the differences between the reference and the estimates of FHR in our method lie within a smaller range around zero, i.e., (-1.9, -0.19), in comparison to [249], i.e., (-8.84, 8.24), implying higher confidence by the proposed method.

## 4.2 Fetal Movement Cancellation in Abdominal ECG Electrocardiogram Recordings Using Signal-to-Signal Translation

As reported by the World Health Organization (WHO), 2.6 million stillbirths occur globally every year which could be avoided by timely access to emergency obstetric care [250]. Proactive monitoring of fetal health conditions allows expectant mothers to be aware of the wellbeing of their babies. Fetal heart rate (FHR) represents an important vital sign to be monitored during intermittent clinical visits in the third trimester. Abnormal FHR

patterns tend to precede adverse prenatal outcomes such as premature birth, hypoxia, or intrauterine growth retardation which may arise in between clinical visits [251]. Continuous monitoring of FHR allows for delivering timely medical interventions to reduce the rate of fetal mortality [251].

In recent years, a variety of technologies have been proposed for FHR monitoring. These include electrocardiogram (ECG) [41], phonocardiogram (PCG) [42], seismo-cardiogram (SCG) [43], and gyro-cardiogram (GCG) [44]. ECG and PCG measure FHR through electrical activity of cardiac muscle and heartbeat sounds respectively, whereas SCG and GCG respectively collect linear and rotational abdominal vibrations. Our observation in [252] has demonstrated that SCG and GCG modalities provide more accurate estimations of FHR when fused with ECG recordings. Fetal ECG (FECG) monitoring could be conducted both invasively and non-invasively [253]. The invasive method aims to acquire FECG data through an electrode attached to the scalp of the fetus during labor [253]. However, this procedure is limited to the period of labor and may induce infections and bruising of the baby's scalp [254]. As an alternative to fetal scalp monitoring, non-invasive fetal ECG (NIFECG) has been investigated recently, showing promise for FHR monitoring [45]. For this purpose, ECG electrodes are placed on the mother's abdomen for data acquisition [46]. However, there are still unexplored challenges complicating FHR extraction through NIFECG, an example of which is the impact of fetal motion on abdominal ECG recordings [255]. This undesired physiological event poses difficulties for estimating fetal ECG components which hold the same level of energy as motion components, as demonstrated in [255]. Hence, the detection and mitigation of fetal

motion artifacts in abdominal recordings would allow for more robust non-invasive FHR estimation.

Conventional methods of non-invasive fetal ECG extraction primarily rely on three stages, namely signal pre-processing, maternal ECG estimation and subtraction, and fetal ECG extraction. The main focus of these methods is on the last two stages [45], whereas other physiological events such as fetal movement could account for inaccurate FHR estimation [255]. Furthermore, these methods are primarily based on deterministic approaches such as independent component analysis (ICA) and adaptive filters, which are less able to extract fetal movement patterns compared to deep neural networks [96].

In this work, a signal-to-signal translation framework for fetal movement cancellation is developed where the noisy abdominal ECG and its corresponding clean counterpart are employed as inputs to a one-to-one mapping neural network. This network is employed to discover cross-domain relations between fetal movement-contaminated and clean abdominal recordings, reinforcing the appearance of clean abdominal ECG signals. The proposed framework leverages generative adversarial networks (GANs) which are widely applicable in computer vision, and adapts them to 1-dimensional signal processing. To the best of our knowledge, this is the first study addressing the cancellation of fetal movement in abdominal ECG recordings through deep neural networks.

## 4.3 Methodology

In the following sub-sections, the dataset, signal-to-signal translation network, and the training procedure are described in detail.

### 4.3.1 Dataset

In this work, we use the fetal ECG synthetic database (FECGGSYNDB) which is publicly available on PhysioNet [256]. This dataset includes 32-channel artificial non-invasive abdominal ECG (AECG) recordings of 10 pregnancies, amounting to 145.8 hours of data. A distinctive characteristic of FECGSYNDB is the inclusion of the ground-truth signals corresponding to fetal ECG (FECG), maternal ECG (MECG), and non-stationary physiological events such as fetal movement and uterine contractions. Motion components provided by FECGSYNDB allow for simulating the non-stationary dynamics of fetal movement in realistic pregnancies. Each recording consists of 5 minutes of data sampled at 250 Hz. In this study, we used the data of 4 channels (27, 28, 29, and 30) which are located below the umbilicus according to [255]. In order to train our model, we need a clean AECG dataset and its corresponding fetal movement-contaminated version. To generate the clean AECG dataset, we combine only maternal ECG and fetal ECG signals. For the noisy version however, maternal ECG, fetal ECG, and fetal movement components are incorporated.

### 4.3.2 Data Pre-Processing

As we employ a deep neural network for motion cancellation in an end-to-end manner, signal preparation merely consists of signal segmentation. As such, each channel of abdominal ECG recordings is segmented into 4-second overlapping windows with 90% overlap between consecutive segments. Hence, each segment represents 1,000 samples (4 seconds at 250 Hz) of data for fetal motion cancellation. These segments are then normalized into a standard distribution (zero-mean and unit variance) to ensure stable convergence when training the network. It is to be noted that signal preparation is applied to both fetal contaminated-AECG and the corresponding clean ground-truth components for supervised training.

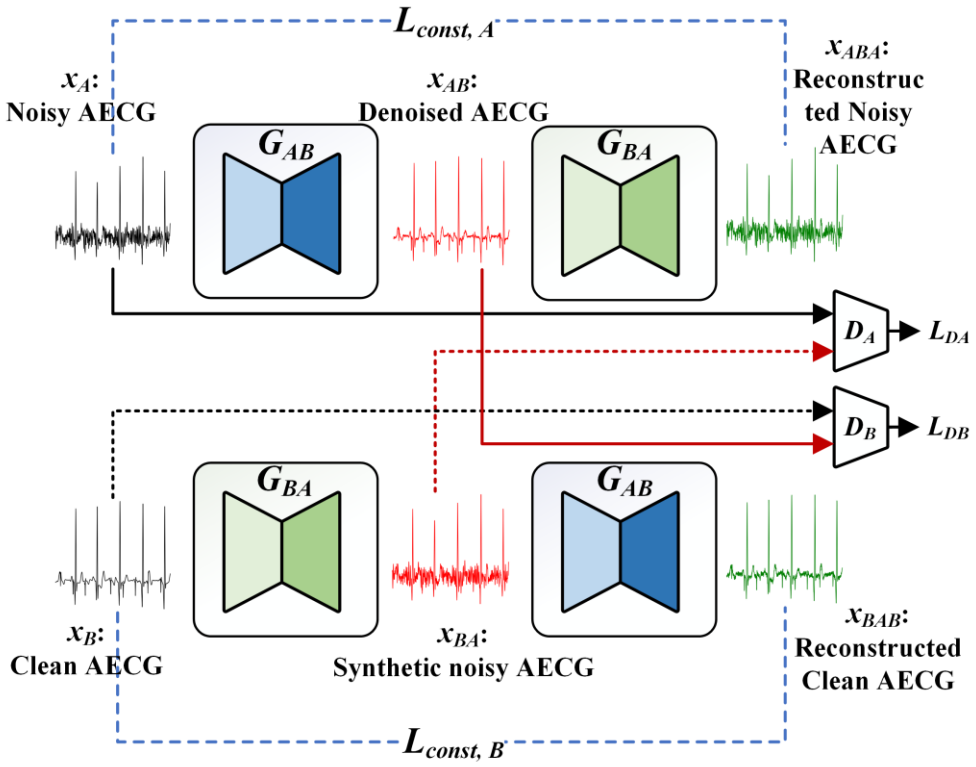


Fig. 4.4 The generative model for discovering cross-domain relations, including two generators ($\boldsymbol{G_{AB}}$ and $\boldsymbol{G_{BA}}$ for noisy AECG-to-denoised AECG and clean AECG-to-noisy AECG translations respectively) and two discriminators ($\boldsymbol{D_A}$ and $\boldsymbol{D_B}$ for scoring synthetic signals generated by $\boldsymbol{G_{BA}}$ and $\boldsymbol{G_{AB}}$ respectively).

### 4.3.3 Signal-to-Signal Translation Network

Fig. 4.4 depicts our developed 1-dimensional (1-D) generative model for fetal movement cancellation. The idea of this network was adopted from DiscoGAN proposed for image cross-domain translations [257], where a synthetic image is generated in one domain given attributes of an image in the other domain. We implemented a modified version of DiscoGAN for 1-D signals to translate noisy abdominal ECG ($x_A$) to its clean counterpart ($x_{BA}$) through generator $\boldsymbol{G_{AB}}$, and similarly a clean AECG ($x_B$) to its motion-contaminated counterpart ($x_{AB}$) using generator $\boldsymbol{G_{BA}}$. The architecture of the generator networks is illustrated in Fig. 4.5. This network is a three-level maxpooling/upsampling U-

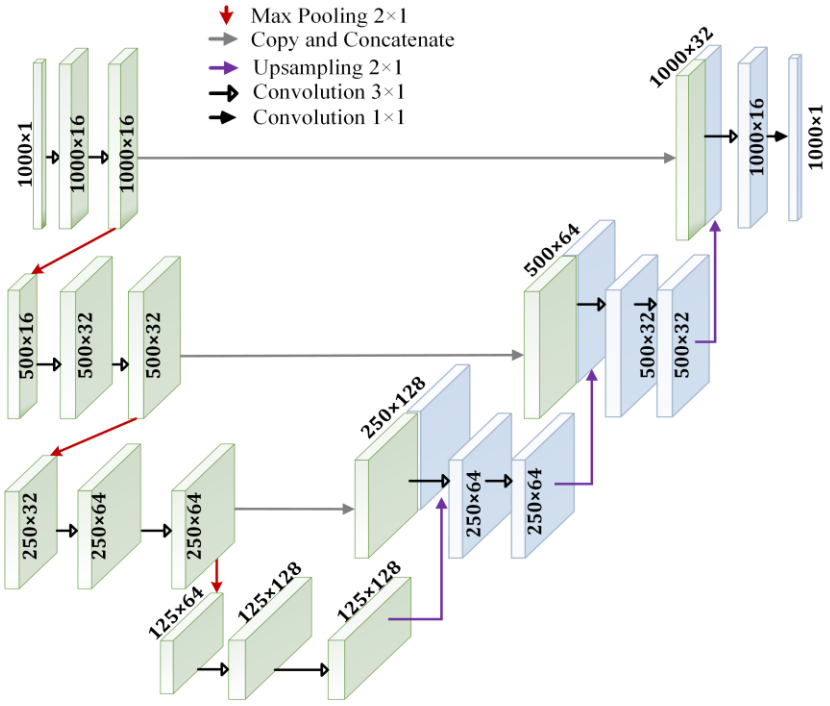


Fig. 4.5 The architecture of the generators adopted for one-to-one mapping in signal-to-signal translation network. This architecture is a 1-D U-Net with three levels of max-pooling and upsampling.

Net with rectified linear unit (ReLU) as the activation function, consisting of an encoder to transform the input signals to an embedding space and a decoder to reconstruct the desired output from the embedded features.

In order to have a one-to-one correspondence between the two domains, we constrain the relation by assuming $\boldsymbol{G_{AB}}$ as the inverse mapping of $\boldsymbol{G_{BA}}$. As such, we expect $\boldsymbol{G_{BA}}$ to be able to reconstruct $x_A$ from $x_{AB}$, which is called $x_{ABA}$. The same assumption stands true for $x_B$. As such,

$$x_{ABA} = \boldsymbol{G_{BA}}(\boldsymbol{G_{AB}}(x_A)),$$
$$x_{BAB} = \boldsymbol{G_{AB}}(\boldsymbol{G_{BA}}(x_B)), \quad (5)$$

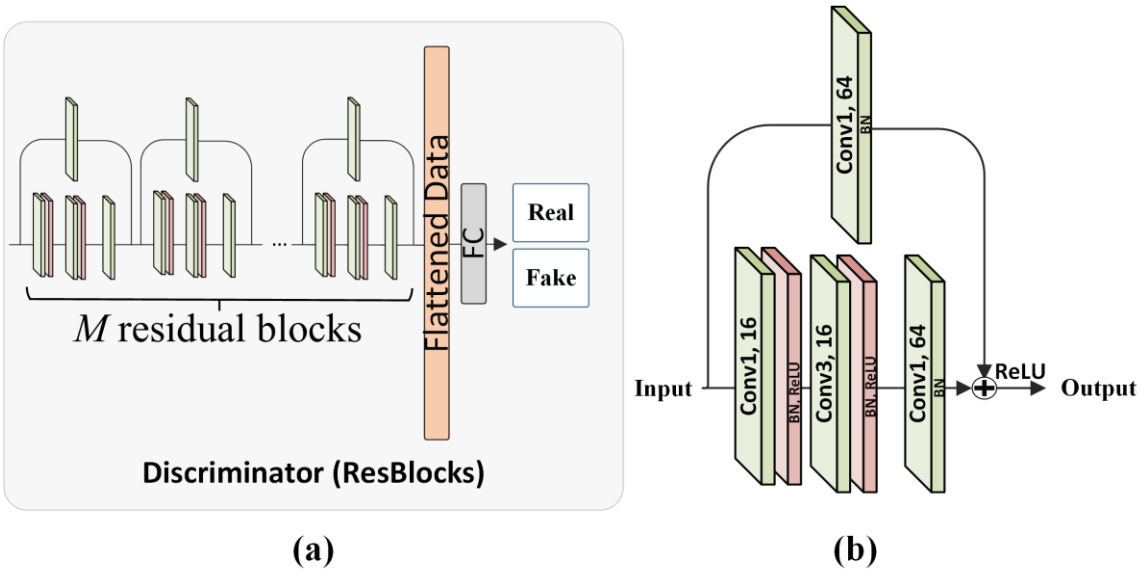


Fig. 4.6 (a) The residual discriminator network for scoring synthetic samples. (b) the residual block including three convolutional layers and a convolutional layer on the skip connection. ReLU and batch normalization (BN) are used to add non-linearity and stable convergence.

which implies the two generative networks, $\boldsymbol{G_{AB}}$ and $\boldsymbol{G_{BA}}$, are coupled to ensure our generative functions can map each domain to its counterpart. These two constraints can be related to the objective function using the L1 distance between the ground-truth signal and its reconstructed version. Therefore, our generative objective function to be minimized during the training is:

$$L_{G,const} = L_{const,A} + L_{const,B} = \\ \|x_{ABA} - x_A\| + \|x_{BAB} - x_B\|, \tag{6}$$

where $L_{G,const}$, $L_{const,A}$, and $L_{const,B}$ represent the total reconstruction loss of the generator, reconstruction loss of $x_A$, and reconstruction loss of $x_B$ respectively. To guarantee inter-domain mapping, each reconstruction channel shown in Fig. 4.4 is scored

with a discriminator, the structure of which is depicted in Fig. 4.6 (a). The discriminator consists of 5 residual layers (shown in Fig. 4.6 (b)) followed by a linear layer using the sigmoid function to generate the score within the range of (0,1). As such, the generated signals $x_{AB}$ and $x_{BA}$ are input to the discriminators of the clean AECG domain ($\boldsymbol{D_B}$) and the noisy ECG domain ($\boldsymbol{D_A}$) respectively, on an adversarial basis. This creates an adversarial loss for the generator networks as below:

$$L_{G,adv} = L_{adv,A} + L_{adv,B} =$$

$$-\mathbb{E}_{x_A \sim P_A}\left[\log \boldsymbol{D_B}\left(\boldsymbol{G_{AB}}(x_A)\right)\right] \qquad (7)$$

$$-\mathbb{E}_{x_B \sim P_B}[\log \boldsymbol{D_A}\left(\boldsymbol{G_{BA}}(x_B)\right)],$$

where $L_{G,adv}$ denotes the adversarial loss which encourages the generators to learn the distribution of data. As a result, the total loss for the generator is the weighted sum of the reconstruction loss and adversarial loss:

$$L_G = \lambda L_{G,const} + (1-\lambda) L_{G,adv}, \qquad (8)$$

where $\lambda$ was set to 0.35 based on our observation.

The adversarial training requires discriminators to distinguish between real and fake distributions. As such, $\boldsymbol{D_A}$ and $\boldsymbol{D_B}$ attempt to learn the distribution of real data more accurately to avoid classifying synthetic data as real, or real data as fake. Hence, discrimination loss for $\boldsymbol{D_A}$ and $\boldsymbol{D_B}$ are defined as

$$L_{D_B} = -\mathbb{E}_{x_B \sim P_B}[\log \boldsymbol{D_B}(x_B)] \\ -\mathbb{E}_{x_A \sim P_A}\left[\log\left(1 - \boldsymbol{D_B}\left(\boldsymbol{G_{AB}}(x_A)\right)\right)\right], \tag{9}$$

and

$$L_{D_A} = -\mathbb{E}_{x_A \sim P_A}[\log \boldsymbol{D_A}(x_A)] \\ -\mathbb{E}_{x_B \sim P_B}\left[\log\left(1 - \boldsymbol{D_A}\left(\boldsymbol{G_{BA}}(x_B)\right)\right)\right], \tag{10}$$

respectively. The sum of the terms in (9) and (10) builds the total adversarial loss of discriminators ($L_D$) as follows:

$$L_D = L_{D_A} + L_{D_B}. \tag{11}$$

The training procedure continues until $\boldsymbol{G_{AB}}$ learns the mapping function from domain $\boldsymbol{A}$ to domain $\boldsymbol{B}$.

### 4.3.4 Training Procedure

The signal-to-signal translation network represents a supervised learning procedure. It thus requires access to raw abdominal ECG recordings and the corresponding clean signals. The training procedure is conducted based on subject-held-out cross-validation. As such, the data of 9 subjects are used for training and the performance is evaluated on the held-out subject. We train our model for 80 epochs on abdominal ECG segments. The Adam optimization algorithm [147] is used with an initial learning rate of 0.001, which is reduced by a factor of 0.98 for every five epochs without performance improvement. The model is implemented on an NVIDIA GeForce RTX 2070 with a batch size of 8.

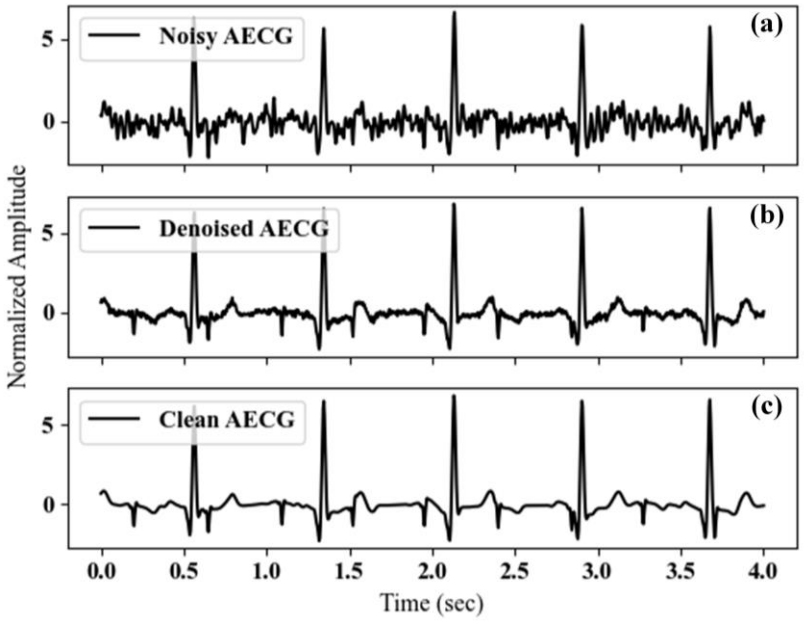


Fig. 4.7 (a) Input noisy abdominal ECG. (b) Denoised abdominal ECG. (c) Clean abdominal ECG.

## 4.4 Experimental Results and Discussion

In this section, the evaluation metrics are introduced, based on which the experimental results are presented.

### 4.4.1 Evaluation Metrics

In this part, we report the performance of the proposed signal-to-signal translation framework using root-mean-square-deviation (RMSD), mean-absolute-error (MAE), and Pearson correlation coefficient (PCC) between the denoised AECG and the input raw AECG signals. Furthermore, RMSD is used to evaluate the robustness to input signal-to-noise ratio (SNR) values. In this work, SNR is defined as the ratio of the power of fetal ECG components to that of fetal movement artifacts. For this, we re-scaled the constituent parts of the abdominal ECG signals to simulate AECG signals at SNR values within the range of (-3, 3) dB.

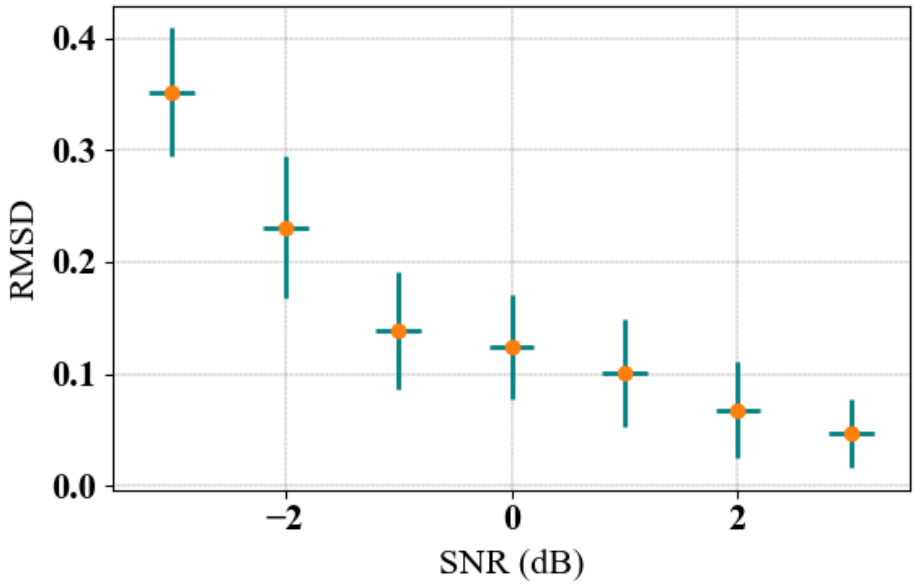


Fig. 4.8 Root-mean-square deviation (RMSD) of the estimated signal in terms of input signal-to-noise ratio (SNR).

### 4.4.2 Performance Evaluation and Discussion

To evaluate the performance of the proposed method, signal segments of the held-out subject were initially normalized and then fed into the network. Fig. 4.7 shows an example of a denoised signal with the input SNR of 0 dB. As can be seen in Fig. 4.7 (b), fetal QRS complexes are visible after the cancellation of fetal movement in the noisy AECG shown in Fig. 4.7 (a). Comparing Figs 4.7 (b) and (c), it is visually demonstrated that the network could mitigate the impact of noise on the AECG signal. Table 4.3 summarizes the performance of our signal-to-signal translation framework in terms of MAE, RMSD, and PCC ($\boldsymbol{\rho_{B,AB}}$(%)) at an input SNR of 0 dB. According to the table, MAE values vary within the range of 0.040-0.181 with an average (± standard deviation) value of 0.099 (± 0.038). The MAE linearly scores the denoised abdominal ECG signal, whereas RMSD aims at considering larger weights for larger error values. As mentioned in Table

Table 4.3 Performance of the denoising method for Subject-held-out Cross-Validation for SNR= 0 $dB$

| Subject | MAE | RMSD | $\rho_{B,AB}(\%)$ |
|---|---|---|---|
| **1** | 0.083 | 0.104 | 99.47 |
| **2** | 0.122 | 0.152 | 98.86 |
| **3** | 0.079 | 0.100 | 99.51 |
| **4** | 0.104 | 0.129 | 99.17 |
| **5** | 0.040 | 0.053 | 99.87 |
| **6** | 0.181 | 0.226 | 97.56 |
| **7** | 0.142 | 0.179 | 98.42 |
| **8** | 0.063 | 0.080 | 99.67 |
| **9** | 0.102 | 0.129 | 99.18 |
| **10** | 0.077 | 0.097 | 99.52 |
| **Mean ($\pm$ std.)** | **0.099 ($\pm$ 0.038)** | **0.124 ($\pm$ 0.047)** | **99.12 ($\pm$ 0.655)** |

4.3, RMSD values vary within the range of 0.053-0.226. The RMSD value for subject 6 suggests a relatively higher estimation error compared to other subjects. However, the average RMSD (± standard deviation) value was reported by 0.124 (± 0.047), implying high morphological similarity between the ground-truth clean abdominal ECG signal ($x_B$) and its denoised counterpart ($x_{AB}$). As described in Table 4.3, PCC values are reported in percentage. Subject 5 with 99.87% and subject 6 with 97.56% suggest the highest and lowest PCC values. The average PCC (± standard deviation) suggests 99.12% (0.655%) morphological consistency between ground-truth and estimated abdominal ECG segments.

To evaluate the robustness of the method against various input SNR values, the signal-to-signal translation network was trained for input SNR values within the range of (-3, 3) dB (-3 dB SNR implies the energy level of fetal movement components is double that of fetal ECG components). Fig. 4.8 illustrates average RMSD values over all ten

subjects in terms of input SNR. The lowest RMSD was = 0.047 for SNR= 3 dB, whereas the largest error (0.352) was suggested by SNR= -3 dB. Small variations of RMSD in terms of SNR demonstrate the robustness of the signal-to-signal translation approach.

## 4.5 Fetal Electrocardiogram Extraction Using Dual-Path Source Separation of Single-Channel Non-Invasive Abdominal Recordings

Non-invasive fetal ECG (NIFECG) has recently shown promise for FHR monitoring [45]. This technology incorporates ECG electrodes placed on the mother's abdomen to acquire abdominal ECG signals [46]. However, these recordings consist of several components including maternal ECG, fetal ECG, uterine contractions, fetal movements, and noise components which complicate FHR estimation [47].

Blind source separation (BSS) techniques such as principal component analysis (PCA) [258], periodic component analysis ($\pi$CA) [259], and independent component analysis (ICA) [260] were used to separate fetal ECG components from abdominal recordings. BSS methods aim to decompose abdominal ECG signals into statistically independent sources including FECG, MECG, and noise. Despite promising results, BSS methods require the number of ECG sensors to be equal to or greater than the number of independent sources [51].

Adaptive filters signify another means of fetal ECG extraction. These filters leverage a transfer function with variable coefficients trained to remove maternal ECG from abdominal recordings. The least mean squares (LMS) filter and recursive least squares (RLS) filter represent two types of adaptive filtering applicable to fetal ECG extraction [246,261–263]. Authors in [264] adopted an echo state neural network (ESN) to project

chest-recorded MECG signals onto a set of non-orthogonal basis functions, where an RLS algorithm was used to update the weights of the network and estimate the fetal ECG signal with an F1-score of 90.2%. The LMS and RLS filters require either a reference maternal ECG channel that morphologically resembles the abdominal recordings, or several linearly independent channels to reconstruct abdominal recordings. The need for reference sensors limits the use of adaptive filters [265]. Alternatively, Kalman filtering approaches require only an arbitrary maternal ECG reference for fetal ECG enhancement. The performance of this method, however, is limited when fetal and maternal QRS peaks coincide [266]. This drawback was addressed by the extended-state Kalman filters which enable separating coincided maternal and fetal peaks at the expense of high computational complexities [267].

Template subtraction (TS) strategies aim to enhance fetal ECG by mitigating the impact of maternal ECG cycles in abdominal recordings. To this end, a maternal ECG template is obtained by averaging the preceding maternal ECG cycles, which is often followed by one or more of the following techniques for fetal ECG extraction: subspace reconstruction by singular value decomposition (SVD) [245], ICA [268], or PCA [258], temporal subtraction of the maternal template [269], and filtering [270]. TS methods can be applied to both single-channel and multiple-channel recordings. In order to use template subtraction methods however, a few assumptions including independence of fetal and maternal components and the consistency of heartbeat cycles in fetal and maternal ECG signals, are made which do not necessarily hold true for abdominal recordings [271]. Having analyzed the traditional methods, we revisit the estimation of fetal ECG and

consider it as an under-determined problem aiming to extract fetal ECG signals from single-channel abdominal recordings.

### 4.5.1 Single-Channel State-of-the-Art

Adaptive filters have recently been proposed for fetal ECG extraction from single-channel recordings [272–278]. The authors in [274] introduced a convolution layer which aims to find the maternal ECG signal as a reference for an RLS adaptive filter. As such, fetal ECG QRS peaks were obtained with an F1-score of 96.0% by subtracting MECG components from abdominal recordings.

Increasing the number of channels using signal decomposition methods is another strategy proposed for fetal ECG extraction. The authors in [279–281] use empirical mode decomposition (EMD) or SVD to decompose the signal or its spectrogram into multiple channels, which is then followed by independent component analysis (ICA) or non-negative matrix factorization (NMF) for fetal ECG separation. These methods are limited to the assumption that the generated channels are statistically independent. Furthermore. some works use decomposition techniques in conjunction with peak detection and coherent averaging to cancel maternal ECG [282,283]. Recently, [284] proposed to factorize the STFT of the AECG signals using NMF, and extract FECG and MECG by setting threshold levels on the activation matrix for maternal and fetal components. As reported, average F1-scores of 94.80% and 84.0% were achieved for the ADFECGDB and NIFECGC datasets, respectively. In [285], the Shannon energy of abdominal time-frequency representation is used to localize maternal ECG peaks for template subtraction, achieving F1-scores of 98.67% and 99.27% on NIFECGC and ADFECGDB, respectively. The performance of

threshold-based methods, however, can be affected if fetal and maternal QRS-peaks coincide, or the abdominal segment length is not long enough to accurately estimate the maternal ECG template.

In recent years, the use of machine learning techniques has attracted extensive attention for fetal ECG monitoring [286–289]. Authors in [286] leveraged a clustering method to classify singularities from abdominal ECG recordings into noise, fetal QRS, and maternal QRS peaks. This method reported F1-scores of 98.04% and 98.63% for ADFECGDB and NIFECGC datasets, respectively. Furthermore, a residual convolutional encoder-decoder network (RCED-Net) was introduced in [287] to separate fetal ECG components from abdominal recordings, where F1-scores of 94.10% and 93.62% were achieved for ADFECGDB and NIFECGC, respectively. Recently in [290], a domain adaptation method was proposed to map the abdominal ECG recordings to their FECG counterparts using a cycle generative adversarial network (CycleGAN). The authors reported an F1-score of 94.7% for the NIFECGC database. A combination of STFT and generative adversarial networks (GANs) was employed in [289] to directly map abdominal recordings to their fetal ECG counterparts. The performance of this method is limited to F1-scores of 93.02% on NIFECGC and 90.05% on ADFECGDB.

In this study, we introduce a single-channel source separation approach based on the dual-path modeling of abdominal ECG signals. This method is coined dual-path source separation (DPSS) architecture, and leverages long short-term memory (LSTM) cells to discriminate the components of FECG and MECG from each other. In our work, the dual-path recurrent neural network (DPRNN), initially introduced in speech processing [291],

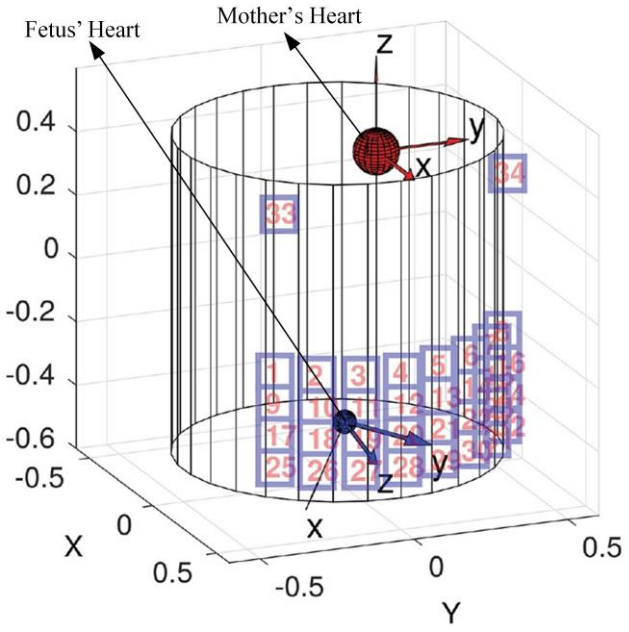


Fig. 4.9 The distribution of ECG sensors on the abdominal region in FECGSYNDB dataset [255].

is utilized to model temporal information of the abdominal ECG signal to denoise and separate fetal ECG components from those of maternal ECG. The effectiveness of the proposed framework is assessed on real and synthetic datasets as explained in the following section.

### 4.5.2 Methodology

#### 4.5.2.1 Experimental Setup & Datasets

Our experimental setup incorporates non-invasive abdominal ECG recordings along with corresponding reference annotation of QRS complexes. To develop our experimental setup, three publicly available datasets including the fetal ECG synthetic database [255,292], the abdominal and direct fetal ECG database [293–295], and the non-invasive fetal ECG challenge database [296] from PhysioNet [176] were used, as described in the following sub-sections.

#### 4.5.2.1.1 Fetal ECG Synthetic Database (FECGSYNDB)

The fetal ECG synthetic database (FECGGSYNDB) includes 32-channel artificial non-invasive abdominal ECG (AECG) recordings of 10 subjects, amounting to 145.8 hours of data. Along with AECG, the ground-truth maternal ECG (MECG) and fetal ECG (FECG) signals are also provided separately. A distinctive characteristic of FECGSYNDB is the inclusion of non-stationary events including fetal movements, uterine contractions, heart rate acceleration/deceleration, and ectopic heartbeats. These non-stationary events ensure simulating the dynamics of realistic pregnancies. Each recording consists of 5 minutes of data sampled at 250 Hz. In this study, we used the data of all 32 channels distributed over the abdominal region as shown in Fig. 4.9. In order to simulate realistic scenarios, pregnancy-specific noise components corresponding to fetal movement, uterine contractions, thermal interference, and maternal heartbeat from [255] are mixed with the fetal ECG signals to generate signal-to-noise ratios (SNR) within the range of (-9, +2) dB. Fetal and maternal acceleration/deceleration scenarios are also incorporated to simulate abnormal scenarios.

#### 4.5.2.1.2 Abdominal and Direct Fetal ECG Database (ADFECGDB)

This dataset comprises 4-channel abdominal ECG recordings sampled at 500 Hz from 10 pregnancy and 12 labor cases between the 38 and 42 weeks of gestation at the Medical University of Silesia. Each labor signal consists of 5 minutes of ECG data acquired from maternal abdomen around the umbilicus (AECG). Additionally, a direct fetal electrocardiogram recorded from the scalp of the fetus (FECG) is provided as the reference signal. Each pregnancy recording includes twenty minutes of abdominal ECG in which

fetal QRS peaks were annotated by an automated fetal QRS detector. Annotations of fetal QRS complexes for both labor and pregnancy datasets were validated by medical experts and provided in the dataset. While the whole dataset is represented by ADFECGDB, labor and pregnancy datasets are shown by ADFECGDB(L) and ADFECGDB(P) in this work, respectively.

#### 4.5.2.1.3 Non-Invasive Fetal ECG Challenge (NIFECGC)

Set A of the non-invasive FECG database from the PhysioNet/computing challenge is the largest publicly available dataset for non-invasive fetal monitoring. This collection consists of 75 4-channel abdominal ECG recordings sampled at 1 kHz with a signal duration of 1 minute. A team of expert clinicians have manually annotated the QRS complexes on the acquired signals. A few recordings (a33, a38, a47, a52, a54, a71, and a74) are excluded from the dataset due to inaccurate QRS annotations as suggested by [297] and [298].

In this study, only the channel with the most evident fetal ECG components was selected per subject in the NIFECGC and ADFECGDB datasets, as summarized in [299]. In single-channel fetal ECG extraction, a signal quality index (SQI) such as that calculated by the naive Bayes classifier in [300] can be computed for the extracted fetal ECG signal to confirm the optimality of the ECG lead position. This mechanism also assists with adjusting the ECG electrodes on the abdominal region in case the signal quality degrades due to fetal movement.

#### 4.5.2.2 Data Preparation

For the sake of uniformity, the NIFECGC and FECGSYNDB datasets were initially resampled to 500 Hz using fast Fourier transform (FFT). Abdominal recordings are often accompanied by baseline wander and power-line noise which should be removed [45]. To remove baseline wander, abdominal ECG signals were high-pass filtered using a zero-phase Butterworth filter with a cut-off frequency of 2 Hz. To attenuate power-line interference, a second-order infinite impulse response (IIR) notch filter centered at 50 Hz or 60 Hz with a quality factor of 25 was applied to the filtered AECG signals in forward and backward directions. Finally, each abdominal signal was segmented into 4-second overlapping windows with 90% overlap between consecutive windows. As such, each

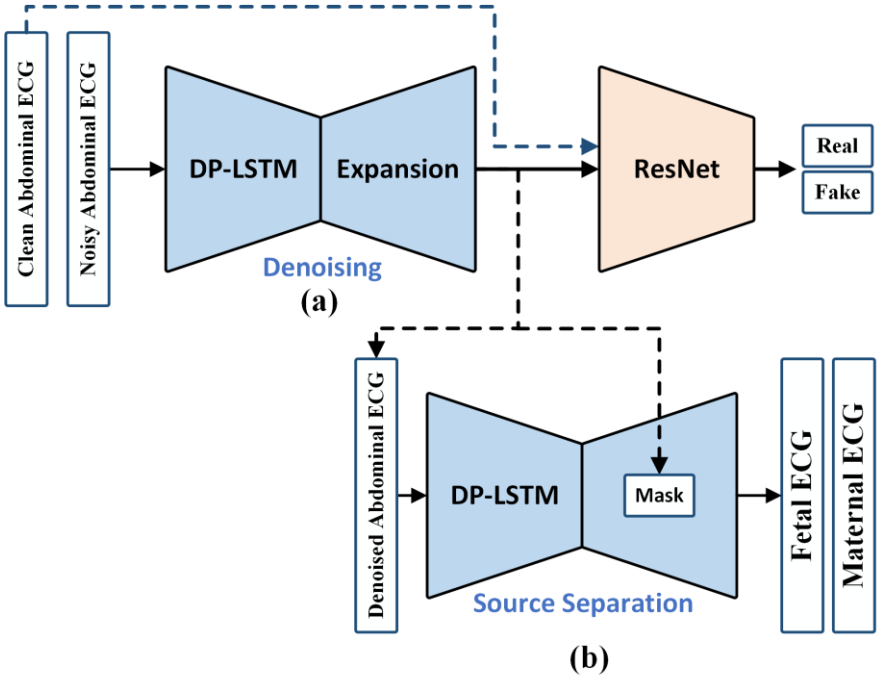


Fig. 4.10 Dual-path source separation (DPSS) framework including (a) a generative network for denoising abdominal ECG, and (b) a dual-path LSTM (DP-LSTM) network for the separation of fetal and maternal ECG signals.

segment represents 2,000 samples (4 seconds at 500 Hz) of data processed for fetal ECG separation. The resulting segments are then normalized into a standard distribution (zero-mean and unit variance) to ensure a stable convergence when training the network.

#### 4.5.2.3 Fetal ECG Extraction

In this section, we describe the proposed approach based on the network illustrated in Fig. 4.10. This network consists of two main blocks, namely denoising and source separation blocks. Initially, the abdominal ECG signal is denoised using a generative network as shown in Fig. 4.10 (a). The denoised abdominal ECG then undergoes a masking-based source separation network as shown in Fig. 4.10 (b). This network extracts the fetal ECG signal by masking maternal components and the maternal ECG signal by masking fetal components. As the dual-path long short-term memory mechanism is used in many parts of the architecture, we first describe this module, and then explain the denoising and source separation networks.

##### 4.5.2.3.1 Dual-Path LSTM Architecture

Long short-term memory (LSTM) networks have been useful for learning sequential data such as speech signals, text, and time series [301]. However, they fail to model long sequences due to vanishing gradients [302]. To address this issue in this work, the dual-path LSTM (DP-LSTM) network is proposed for modeling abdominal ECG signals as shown in Fig. 4.11. The intuition behind the DP-LSTM is modeling shorter segments (local modeling), rather than the whole signal, and then aggregating the information across all segments (global modeling). For this purpose, a sequential input $\boldsymbol{I}_l \in \mathbb{R}^{M \times T}$ with $M$ and $T$ representing the number of input channels and timesteps respectively,

is segmented into $K$ segments of length $L$ ($T = K \times L$). As such, a segmented feature map $\mathbf{S_l} \in \mathbb{R}^{M \times K \times L}$ is obtained.

Fig. 4.11 (a) illustrates the local modeling path. In this path, a bi-directional LSTM (Bi-LSTM) module is employed to model each sequence forward and backward to capture past and future information, respectively. The output of this layer represents a matrix with a different number of channels, i.e., $\mathbf{B}_l \in \mathbb{R}^{H \times K \times L}$. This difference occurs as the input and output vectors in an LSTM block are not necessarily of the same dimensions. To restore the original dimensions ($M \times K \times L$), a fully-connected (FC) layer is then applied to $\mathbf{B}_l$, which results in $\mathbf{F_l} \in \mathbb{R}^{M \times K \times L}$. In order to accelerate the learning process and enhance the generalizability of the model [161], a normalization layer is used as below:

$$\mathbf{N}_l = \frac{\mathbf{F}_l - \mathbb{E}\{\mathbf{F}_l\}}{\sqrt{\mathbf{Var}\{\mathbf{F}_l\} + \epsilon}} \odot \gamma + \beta, \qquad (12)$$

where $\mathbf{N}_l \in \mathbb{R}^{M \times K \times L}$, $\gamma \in \mathbb{R}^{M \times 1}$, $\beta \in \mathbb{R}^{M \times 1}$, and $\epsilon$ represent the normalized matrix, normalization scaler, normalization bias, and numerical stability value respectively. In (12), $\mathbb{E}\{.\}$, $\mathbf{Var}\{.\}$, and $\odot$ denote expectation value, variance, and the Hadamard product operator respectively. The input feature map is transferred to the output layer through a skip connection, allowing for information flow from the initial layer to the last layer. As such, the output of the local modeling path, shown by $\mathbf{O_l}$, is a summation of the normalized matrix and input feature map ($\mathbf{N}_l$ and $\mathbf{S_l}$ respectively), which serves as the input to the global modeling path.

Fig. 4.11 (b) illustrates the global modeling path in the DP-LSTM network. The structure of this path is similar to that of the local modeling path except for the direction of

the Bi-LSTM module which is applied to each of the $L$ sequences separately. The output of the Bi-LSTM, fully-connected, and normalization layers are shown by $\mathbf{B}_g$, $\mathbf{F}_g$, and $\mathbf{N}_g$ respectively. The output of this layer is a summation of the input feature and the normalized matrix ($\mathbf{N}_g$ and $\mathbf{S}_g$ respectively). As a result of the local and global modeling paths, the information of the input signal is extracted for further processing.

#### 4.5.2.3.2 Abdominal ECG Denoising

As mentioned earlier, abdominal recordings are contaminated by noise components. Generative adversarial networks (GAN), initially proposed by Ian Goodfellow et. al [303], have been widely used in a variety of applications, an instance of which is ECG signal denoising [304], biomedical image synthesis [136], and speech

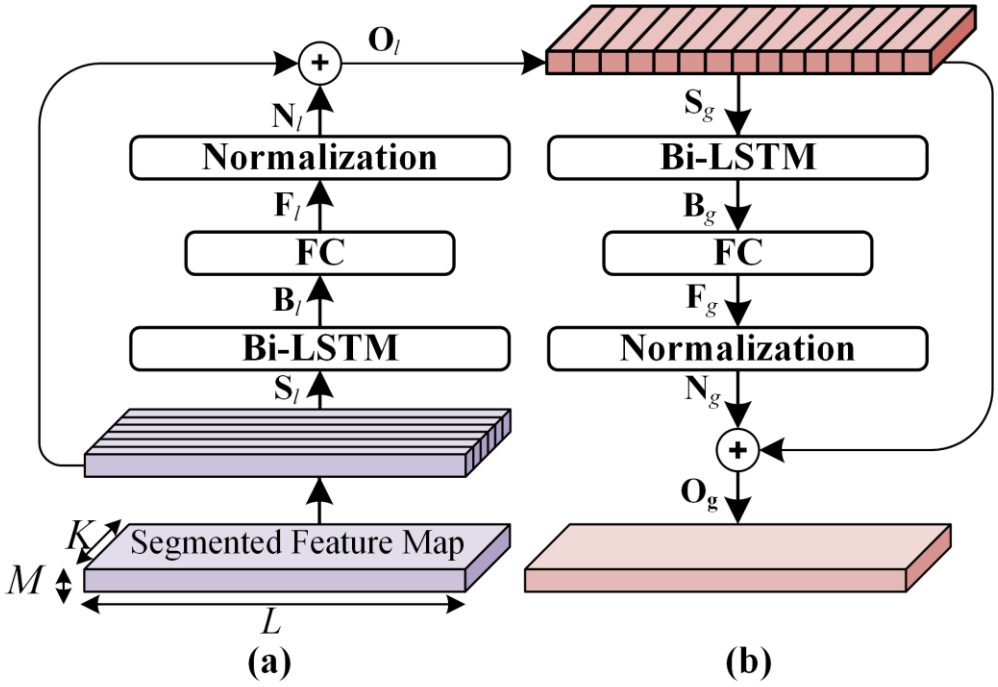


Fig. 4.11 Dual-path LSTM architecture for sequential modeling. The network includes two modeling paths, (a) intra-segment (local: path 1) and (b) inter-segment (global: path 2).

synthesis [305]. Fig. 4.12 shows the developed conditional least-square generative adversarial network (LSCGAN) for noise mitigation. This network is composed of two sub-modules, a DP-LSTM generator and a discriminator based on residual blocks, as illustrated in Fig. 4.12 (a) and described below:

**Generator:** As shown in Fig. 4.12 (a), the generator consists of a 1-D convolutional layer with 16 filters, a DP-LSTM module, and a depth-wise convolution. The 1-D convolutional layer takes a single-channel noisy abdominal ECG signal $x_c \in \mathbb{R}^{1\times T}$, and transforms it into $x_c \in \mathbb{R}^{M\times T}$, where $M = 16$ and $T$ shows the length of the signal. The DP-LSTM block then segments the signal and extracts its temporal features ($\mathbf{O_l} \in \mathbb{R}^{M\times K\times L}$), which are then expanded back to their original dimensions ($\mathbb{R}^{M\times T}$). The resulting $M$ channels are transformed to a denoised AECG signal, i.e., $G(x_c)$, using a depth-wise convolution (Conv3, 1).

**Discriminator:** The discriminator consists of $m$ 1-D residual blocks (ResBlocks). Each residual block incorporates three convolutional layers accompanied by a skip connection as shown in Fig. 4.12 (b), where the rectified linear unit (ReLU) and batch normalization (BN) are employed to add non-linearity and stabilize the training procedure, respectively. The ResBlocks are followed by a fully-connected (FC) layer to discriminate between real ($x$) and synthetic data ($G(x_c)$).

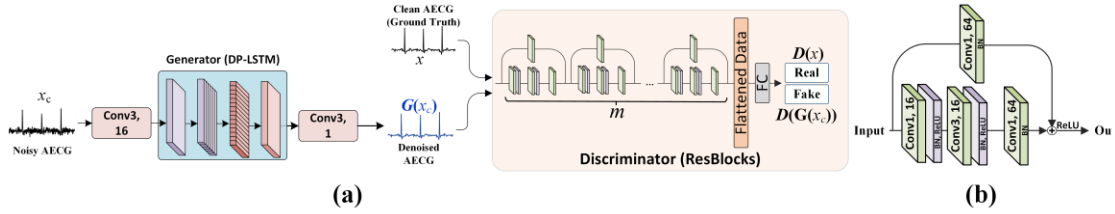


Fig. 4.12 (a) The CGAN-based denoising network consisting of a dual-path LSTM (DP-LSTM) generator and a discriminator based upon residual blocks. (b) A residual block structure.

**Training:** In GAN architectures, the generator and discriminator networks are trained simultaneously. Given that $x_c$ and $x$ are sampled from the noisy abdominal ECG distribution ($p_{data}(x_c)$) and the clean abdominal ECG distribution ($p_{data}(x)$), the generator aims to map $x_c$ to $x$ through adversarial training. On the other hand, the discriminator is supposed to classify samples from $p_{data}(x)$ as real and samples from $p_{data}(x_c)$, i.e., $G(x_c)$, as fake. As such, the generator tries to imitate the training data and deceive the discriminator, whereas the discriminator is trained to distinguish fake data. The training procedure is conducted in two steps. In the first step, the discriminator is trained to classify $x$ as real data ($p_{data}(x)$). In the second step, the generator learns to imitate $p_{data}(x)$ such that the discriminator classifies $G(x_c)$ as real data. Hence, two loss terms corresponding to the discriminator and generator networks are defined as (13) and (14), respectively.

$$\min_D L_{\text{LSCGAN}}(D) = \frac{1}{2}\mathbb{E}_{x\sim p_{data}(x)}[(D(x)-1)^2] + \frac{1}{2}\mathbb{E}_{x_c\sim p_{data}(x_c)}[D(G(x_c))^2], \quad (13)$$

$$\min_G L_{\text{LSCGAN}}(G) = \frac{1}{2}\mathbb{E}_{x_c \sim p_{data}(x_c)}\left[\left(D\left(G(x_c)\right) - 1\right)^2\right] + \lambda\|G(x_c) - x\|_1. \quad (14)$$

In (14), the term $\lambda\|G(x_c) - x\|_1$ is used to minimize the distance (L1-norm) between $G(x_c)$ and $x$. This term is fused with the adversarial loss term with a coefficient of $\lambda$ which was found to be 0.01 based on our empirical results. The whole process of training repeats until the generator learns the distribution of clean abdominal ECG signals. Therefore, the generator would be capable of denoising abdominal ECG signals.

#### 4.5.2.3.3 Mask-Based Source Separation

Mask-based source separation is illustrated in Fig. 4.13. This step receives the denoised abdominal ECG signal from the denoising network, extracts features, and eventually separates fetal and maternal components from each other.

**Sparse Feature Extraction:** ECG signals hold sparse nature composed of regular activities (P, QRS, and T waveforms) as well as periods of inactivity [69]. To assure accurate source separation, we need to extract fine and coarse features of abdominal ECG signals. To this end, we utilize an inception module with dimensionality reduction as codenamed by [70] for the GoogleNet architecture and depicted in Fig. 4.13 (a). As such, the denoised abdominal ECG of length $T$ is segmented into a matrix of dimensions $K \times 1 \times L$. This matrix undergoes the inception module with four paths. These paths leverage convolutional layers with filter sizes of 1, 3, and 5 as well as a pooling layer used to reinforce invariance to signal distortion. The number of paths (the limitation on filter sizes) will be investigated in the following sections. The outputs of all paths are then

concatenated to build a feature map of dimensions $K \times M \times L$. The logic behind such feature extraction is to transfer the abdominal ECG signal into a higher level of abstraction to simplify the separation procedure. The output of the inception module serves as the input to the dual-path masking block as illustrated in Fig. 4.13 (b).

**Dual-path Masking:** To separate fetal and maternal features from each other, their respective features are masked in high levels of abstraction. To this end, a series of DP-LSTM modules are cascaded to extract masking functions from the sparse feature map as shown in Fig. 4.13 (b). The output of the DP-LSTM blocks is a matrix with dimension $K \times M \times L$. To extract maternal and fetal components, two masking functions are required to be multiplied by the sparse feature map which is of dimension $K \times M \times L$. These masking

functions are built from the output of DP-LSTM series using a 2D convolutional layer with $2M$ filters followed by a sigmoid function. As a result, a masking function with a size of $K \times 2M \times L$ is obtained which is then split into two masking maps of sizes $K \times M \times L$. Next, each masking function is multiplied by the sparse feature map separately to discriminate features representing fetal and maternal components. Each of the acquired

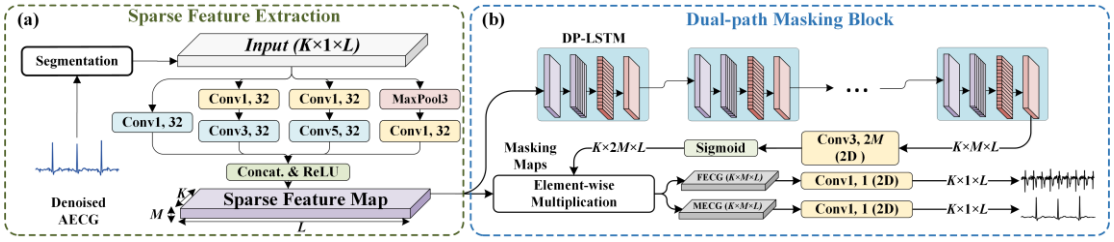


Fig. 4.13 Source separation network including (a) sparse feature extraction and (b) dual-path masking blocks. The input to the network is a denoised abdominal ECG which undergoes feature extraction and masking. The fetal and maternal ECG signals constitute the output of the network.

FECG and MECG components consist of $M$ features along the second dimension. To transform the features back into the time-domain, the $M$ features should be consolidated into a single feature. As such, two 2-D depth-wise convolutional layers using a single feature map are utilized to generate FECG and MECG signal segments. Finally, the segments are assembled to build the original FECG and MECG signals.

**Training:** The objective of training the source separation network is to maximize the signal-to-noise ratio (SNR). As the temporal variation of the signal, rather than its order of magnitude, holds information about the cardiac cycles, we optimize the scale-invariant SNR (SI-SNR) function which is robust to the magnitude scales of the estimated signal [71]. For a ground-truth signal x and its estimated counterpart $\hat{x}$, SI-SNR is defined as follows:

$$\text{SI-SNR} = 10 \log \frac{\|x_{\text{target}}\|^2}{\|e_{\text{noise}}\|^2}, \tag{15}$$

where

$$x_{\text{target}} = \frac{\langle \hat{x}, x \rangle x}{\|x\|^2}, e_{noise} = \hat{x} - x_{\text{target}}, \tag{16}$$

and $\|.\|$ and $\langle . \rangle$ denote the Euclidean norm and scalar product respectively. As such, the loss function for training the source separation network is defined as the negative value of the SI-SNR, as mentioned below:

$$\min L_{\text{SI-SNR}} = -10 \log \frac{\|x_{\text{target}}\|^2}{\|e_{\text{noise}}\|^2}. \tag{17}$$

During training, the error between fetal ECG ($s_{\text{FECG}}$) and maternal ECG ($s_{\text{MECG}}$) signals and their estimated counterparts [$\hat{s}_{\text{FECG}}, \hat{s}_{\text{MECG}}$] are achieved using the $L_{\text{SI-SNR}}$ loss

function. However, since the model estimates the masking maps simultaneously using the same input abdominal ECG, the order of the output signals is unknown in advance. As such, the output can have either $[\hat{s}_{\mathrm{FECG}}, \hat{s}_{\mathrm{MECG}}]$ or $[\hat{s}_{\mathrm{MECG}}, \hat{s}_{\mathrm{FECG}}]$ arrangements. This issue causes training failure. To address this problem, we employ the permutation invariant training (PIT) strategy [306], where the two possible permutations for FECG and MECG are considered. As such, at every training iteration, the average errors for both orders are

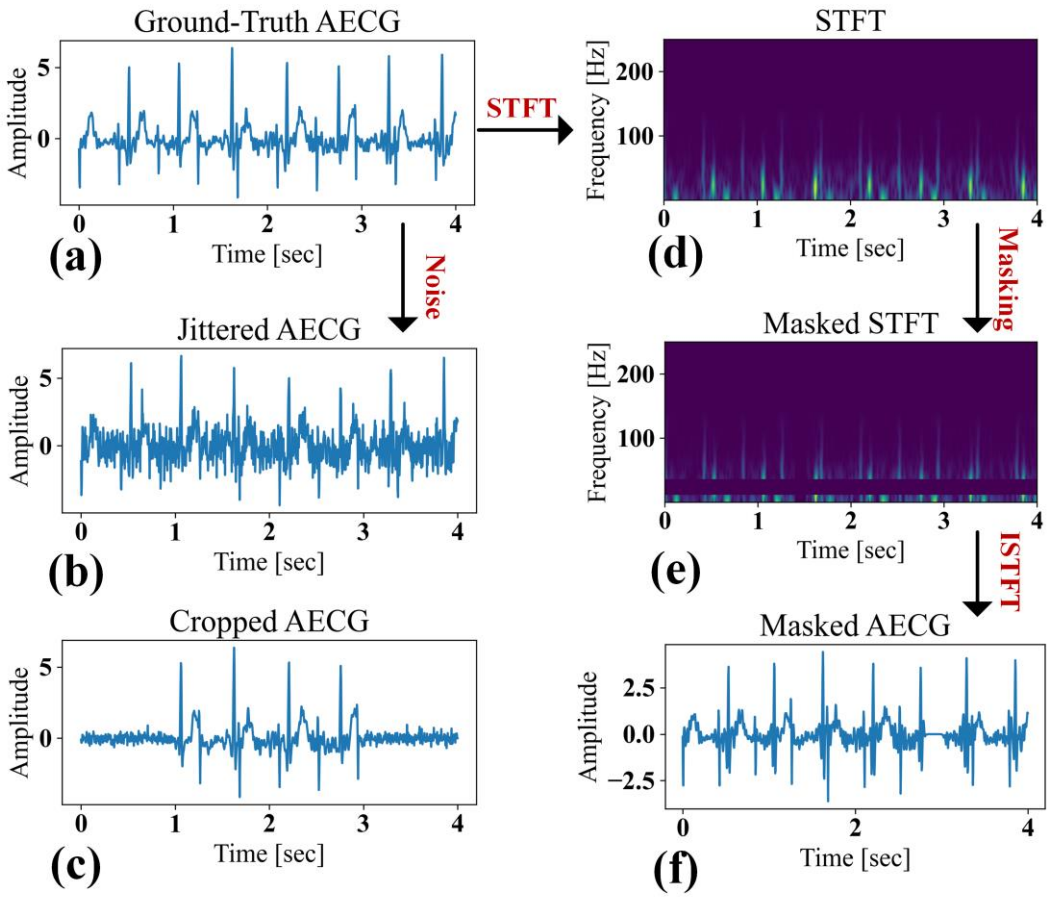


Fig. 4.14 Data augmentation techniques applied on (a) abdominal ECG (AECG) from FECGSYNDB. (b) jittered AECG, (c) cropped AECG, (d) STFT of AECG, (e) masked STFT of AECG, and (f) masked ECG.

evaluated, and the order with the least average loss is selected to back-propagate the error and update the weights of the network.

#### 4.5.2.4 Data Augmentation

Neural networks tend to suffer from overfitting due to the limited amount of training data [307]. This phenomenon leads the network to excessively fit the training data, restraining the generalizability of the network to unseen data [308]. On the other hand, neural networks can benefit from data augmentation which aims at generating synthetic data by transforming existing samples into their modified versions [309]. The augmented data space enforces stability in training by reducing the variance [309]. In this study, we employ three techniques, namely jittering, cropping, and time-frequency masking, to augment the training dataset. Fig. 4.14 shows the ground-truth signal and the augmented samples.

##### 4.5.2.4.1 Jittering

An effective data augmentation method is jittering wherein zero-mean white Gaussian noise, i.e., $\mathcal{N}(0, \sigma^2)$, is added to training samples. Furthermore, pregnancy-specific noise such as uterine contractions and fetal movement are combined with abdominal signals to simulate realistic scenarios. An example of jittering is illustrated in Fig. 4.14 (b), where an abdominal ECG from FECGSYNDB is contaminated by noise components. This practice helps mitigate the time series drift caused by unseen data [310]. In this study, the standard deviation of the normal distribution is selected such that the noise power falls 3 dB below the energy of FECG QRS complexes.

#### 4.5.2.4.2 Cropping

Random cropping aims to create a random subset of original abdominal ECG recordings by cropping a random section of the signal. QRS complexes in an abdominal recording are not always wholly visible in the signal segment which might hinder the network to learn such cases. Random cropping deals with these cases, helping the network to learn the data distribution more comprehensively, and enhance generalizability. Hence, the Beta distribution whose superiority for data augmentation over the uniform distribution has been demonstrated [311], was utilized to specify the cropping width, $w$. This value is selected as below:

$$w = \lfloor w'T \rfloor, w' \sim \mathbf{Beta}(\beta, \beta), \qquad (18)$$

where $\beta \in (0, +\infty)$, and $\lfloor . \rfloor$ and $T$ denote the floor function and the original segment length, respectively. As such, a window of length $w$ centered in the middle of the original segment is cropped and the held-out samples are padded using a zero-mean normal distribution with SNR=10 dB. An example of a 50% random cropping is presented in Fig. 4.14 (c), where a fetal QRS peak is located on the boundary.

#### 4.5.2.4.3 Time-Frequency Masking

In this technique, the time-frequency representation of each segment is obtained through the short-time Fourier transform (STFT) as shown in Fig. 4.14 (d). Then, a number of randomly selected bins are set to zero, which builds a new time-frequency representation as illustrated in Fig. 4.14 (e). Finally, the inverse STFT (ISTFT) is applied to the generated time-frequency representation to transform it back into the time domain. This technique can attenuate certain spectral features in ECG (such as T-waves as shown in Fig. 4.14 (f))

as well as the temporal evolution of the signal (e.g., around 2.9 s in Fig. 4.14 (f)). As such, time-frequency masking reinforces the network with learning a more general distribution. It should be noted that random bins are chosen from frequency bins below 50 Hz due to the concentration of ECG information in this region [312].

### 4.5.3 Experimental Results

In this section, training procedures and experimental results are discussed. Moreover, an ablation study is presented to investigate the network configuration.

#### 4.5.3.1 Training Procedure

The dual-path source separation (DPSS) represents a supervised learning configuration. It thus demands access to noisy abdominal ECG, clean abdominal ECG, fetal ECG, and maternal ECG signals. The only dataset which provides all the required signals is FECGSYNDB, from which the data of ten different pregnancies were used for training. The training procedure was conducted based on leave-subject-out cross-validation. As such, the data of 9 subjects were used for training and the performance was evaluated on the held-out subject. For the sake of computational simplicity, the denoising and source separation networks were trained separately. To further evaluate the performance of the method, the trained model for FECGSYNDB was tested on ADFECGDB(P), ADFECGDB(L), and NIFECGC datasets as well.

We trained our model for 200 epochs on abdominal ECG segments. The Adam optimization algorithm [147] was used with an initial learning rate of 0.001, which was reduced by a factor of 0.98 for every five epochs without performance improvement. The model was implemented on an NVIDIA GeForce RTX 2070 with a batch size of 32.

Fig. 4.15 (a) illustrates a 10-second noisy abdominal ECG segment of recording a02 from NIFECGC. This signal was segmented into overlapping 4-second chunks, denoised, source-separated, and eventually reconstructed by merging the segments. As such, Figs 4.15 (b), (c), (d), and (e) depict the merged output of the denoiser network,

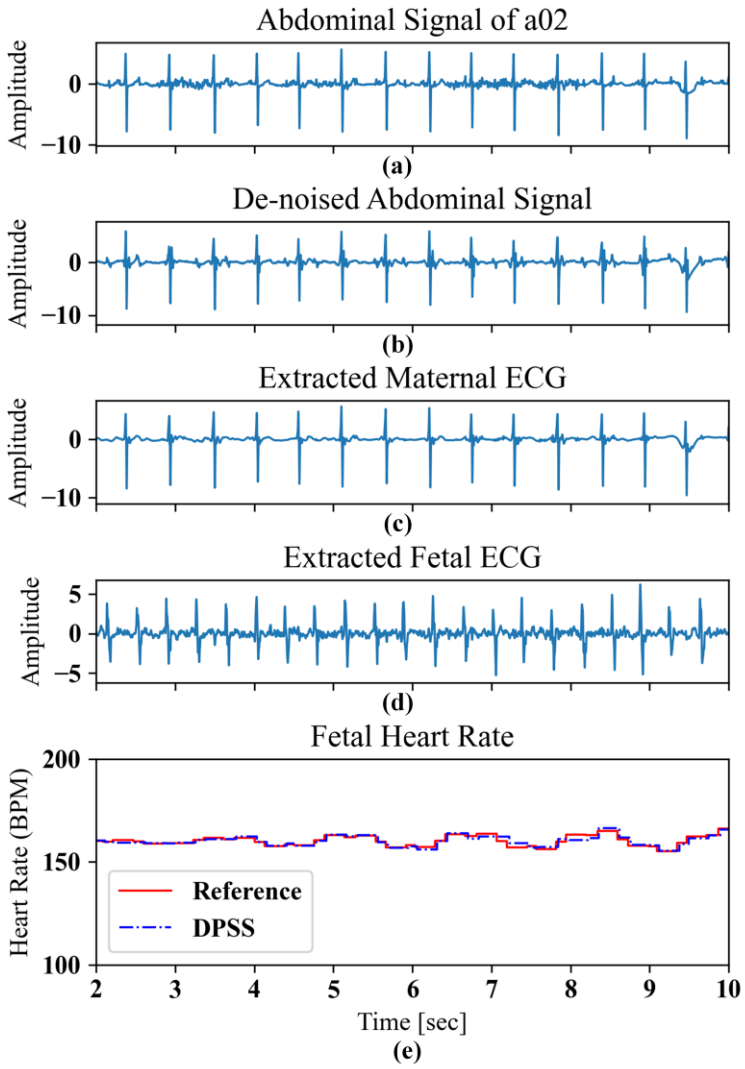


Fig. 4.15 An 8-second test segment reconstructed by overlapping the segments of a02 from the NIFECGC dataset. (a) Noisy abdominal recording, (b) de-noised abdominal ECG, (c) extracted maternal ECG, (d) extracted fetal ECG, and (e) estimated fetal heart rate.

maternal ECG, fetal ECG, and estimated FHR, respectively. These figures visually demonstrate the effectiveness of DPSS in fetal ECG extraction. As observed in Fig. 4.15 (d), the peaks corresponding to maternal QRS waveforms have been majorly attenuated, allowing for the detection of the FECG R-peaks. In parallel, Fig. 4.15 (e) illustrates high consistency between the estimated FHR values by DPSS with their respective ground-truth values.

#### 4.5.3.2 Evaluation Metrics

##### 4.5.3.2.1 Statistical Analyses of Fetal QRS Detection

Statistical analyses were conducted to evaluate the performance of the DPSS network and compare it with the state-of-the-art methods of single-channel fetal ECG extraction. In order to compare the extracted fetal ECG segments with their ground-truth counterparts, the modified version of the Pan-Tompkins method for fetal QRS detection was employed [255]. Following the detection of R-peaks in both reference and extracted FECG signals, sensitivity (Sen), positive predictive value (PPV), accuracy (Acc), and F1-score (F1) were obtained as follows:

$$\text{Sen} = 100 \times \frac{TP}{TP+FN}, \tag{19}$$

$$\text{PP} = 100 \times \frac{TP}{TP+FP}, \tag{20}$$

$$\text{Acc} = 100 \times \frac{TP}{TP+FP+FN}, \tag{21}$$

$$\mathrm{F1} = \frac{2 \times PP \times Sen}{PP + Sen}, \qquad (22)$$

where *TP*, *FP*, and *FN* represent true positives (correctly detected FECG QRS complexes), false positives (falsely detected non-existent FECG QRS complexes), and false negatives (falsely missed existent FECG QRS complexes), respectively. According to [46], a detected QRS is considered a true positive if it is within 50 ms from the reference R-peak. Additionally, extracted FECG segments are assessed by the FHR precision, which is defined as the percentage of segments wherein the extracted FHR is within $\pm 5$ beats-per-minute (BPM) of the reference FHR.

#### 4.5.3.2.2 Clustering-based Source Separation Analyses

As mentioned earlier, DPSS generates two $K \times M \times L$ masking maps to separate fetal and maternal ECG signals. These masking maps are obtained by learning the distributions of MECG and FECG signals from the input AECG signals. A well-trained network is meant to acquire two separable distributions in the embedding space, corresponding to fetal and maternal ECG components. In this work, we quantify the quality of separation by the Davies-Bouldin index defined below:

$$DB = \frac{\sigma_M + \sigma_F}{2d(c_M, c_F)}, \qquad (23)$$

where $c_M$ and $c_F$ represent the centroids of the maternal and fetal clusters, respectively, and $\sigma_M$ and $\sigma_F$ show the average distance of all maps in maternal and fetal clusters to their respective centroids respectively. Also, $d(c_M, c_F)$ represent the distance between the centroids of maternal and fetal clusters. This metric implies the ratio of intra-cluster

distances to those of inter-clusters. As such, lower values of DB indicate better clustering. In this work, we conduct a two-class K-means algorithm to cluster masking maps. We then find the Davies-Bouldin index using the ground-truth labels (fetal or maternal).

Furthermore, we compute the Jaccard index to assess the similarity between the clustered group and its corresponding labels (either maternal or fetal). The Jaccard index is defined as follows:

$$\text{Jaccard}(A, B) = \left|\frac{A \cap B}{A \cup B}\right| = \frac{tp}{tp+fp+fn}, \qquad (24)$$

where $tp$, $fp$, and $fn$ signify the numbers of truly-included elements, falsely-included elements, and falsely-excluded elements in a cluster, respectively. To compute the Jaccard metric, each cluster is assigned to either maternal or fetal based on the labels of the majority of the samples. The Jaccard metric is the average of the Jaccard values of the maternal and fetal clusters.

The last parameter is the purity index. This parameter describes the extent to which clusters contain their dominant elements. For each cluster, the dominant elements are counted and summed over all clusters. The sum value is then divided by the total number of elements.

#### 4.5.3.3 Model Performance

Table 4.4 summarizes the statistical analysis of the performance of the dual-path source separation (DPSS) framework on the FECGSYNDB dataset. The results in Table 4.4 correspond to a DPSS architecture with three dual-path LSTMs and six feature

Table 4.4 Cross-Validation Performance for FECGSYNDB Dataset

| Subject | Sen. (%) | PPV (%) | Acc (%) | F1 (%) | SI-SNR (dB) |
|---|---|---|---|---|---|
| Subject 1 | 98.78 | 99.77 | 98.55 | 99.27 | 14.62 |
| Subject 2 | 97.43 | 99.07 | 96.55 | 98.24 | 13.35 |
| Subject 3 | 98.95 | 99.07 | 98.04 | 99.01 | 16.91 |
| Subject 4 | 98.82 | 99.54 | 98.37 | 99.18 | 15.36 |
| Subject 5 | 97.54 | 99.21 | 96.79 | 98.37 | 14.32 |
| Subject 6 | 99.02 | 99.60 | 98.62 | 99.31 | 15.02 |
| Subject 7 | 98.61 | 99.71 | 98.32 | 99.16 | 14.67 |
| Subject 8 | 99.09 | 99.61 | 98.71 | 99.35 | 15.94 |
| Subject 9 | 98.65 | 99.83 | 98.48 | 99.24 | 13.88 |
| Subject 10 | 98.58 | 99.76 | 98.35 | 99.17 | 15.13 |
| **Average (std.)** | **98.55 (± 0.59)** | **99.52 (± 0.29)** | **98.08 (± 0.77)** | **99.03 (± 0.39)** | **14.92 (± 1.02)** |

extraction filters with lengths 1, 3, 5, 7, 9, and 11. The length of the sub-segments was set to $L$= 100 ms for temporal modeling in DPSS. Per each held-out subject, the model is trained on the augmented data of 9 other subjects, which includes 3,432,960 4-second segments of AECG signals. The performance is then reported by the sensitivity (Sen), positive predictive value (PPV), accuracy (Acc), F1-score (F1), and the maximum SI-SNR achieved during the training for the 190,720 test segments. As mentioned in the table, the best performance is achieved for subject 8 with sensitivity, positive predictive value, accuracy, F1-score, and SI-SNR of 99.09%, 99.61%, 98.71%, 99.35%, and +15.94 dB respectively. The weakest source separation performance belongs to subject 2 with 98.24% and +13.35 dB F1-score and SI-SNR respectively. However, the average (± standard deviation) F1-score and SI-SNR suggest 99.03% (± 0.39) agreement with the reference

QRS complexes and +14.92 (± 1.02) dB signal-to-noise ratio in the output, respectively. Such a decent performance is expected as the test set is sampled from the same dataset and thus the same data distribution as the training set.

Table 4.5 Performance Evaluation of DPSS on the ADFECGDB and NIFECGC Datasets

| Dataset | FHR precision (%) | Sen (%) | PPV (%) | Acc (%) | F1 (%) |
|---|---|---|---|---|---|
| **ADFECGDB(P)** | 87.88 | 96.50 | 97.98 | 94.67 | 97.24 |
| **ADFECGDB(L)** | 97.43 | 99.07 | 96.55 | 98.24 | 98.08 |
| **ADFECGDB** | 98.95 | 99.07 | 98.04 | 99.01 | 97.70 |
| **NIFECGC** | 98.82 | 99.54 | 98.37 | 99.18 | 95.30 |
| **Average of the two datasets** | 86.23 | 95.75 | 97.29 | 93.43 | 96.50 |

To further evaluate the generalizability of DPSS, the trained model is tested on the ADFECGDB(P), ADFECGDB(G), and NIFECGC datasets with 2,999, 749, and 149 segments per channel respectively to reconstruct FECG signals, and their statistical results are summarized in Table 4.5. According to Table 4.5, the labor dataset achieves the highest F1-score (98.08%) and FHR precision (89.22%), while NIFECGC obtains the weakest performance with 83.86% of FHR precision and 95.3% of F1-score. ADFECGDB, which denotes the average result of labor and pregnancy dataset, suggests 97.7% and 88.61% F1-score and FHR precision, respectively. The slight difference of statistical results of ADFECGDB and NIFECGC can be explained by stronger fetal QRS peaks in ADFECGC dataset. The last row of Table 4.5 signifies the average performance of DPSS on real datasets, i.e., NIFECGC and ADFECGDB, with sensitivity of 86.23%, positive predictive

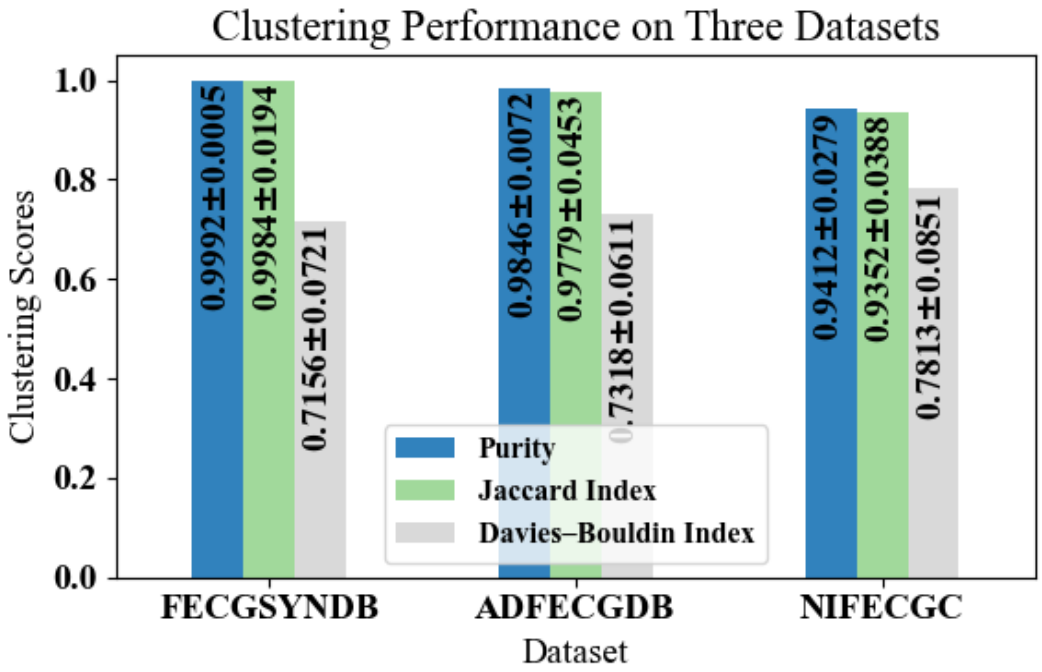


Fig. 4.16 Clustering performance for FECGSYNDB, ADFECGDB, and NIFECGC datasets based on the purity, Jaccard, and Davies-Bouldin indices.

value of 95.75%, accuracy of 97.29%, and F1-score of 93.43%, demonstrating the robustness of the method.

#### 4.5.3.4 Clustering-based Evaluation

As introduced earlier, the purity, Jaccard, and Davies-Bouldin indices are used to quantify and evaluate the quality of source separation in DPSS. As shown in Fig. 4.13, two masking maps (each of size $K \times M \times L$) are produced to separate the fetal and maternal components of a 4-second mixture AECG signal. For ideal source separation, we expect the maternal and fetal masking maps of all segments to form two separate clusters in the $\mathbb{R}^{K\times M\times L}$ space. To investigate this phenomenon, the aforementioned metrics can be leveraged provided that each $K \times M \times L$ masking map represents a sample in the $\mathbb{R}^{K\times M\times L}$

space. However, the computation of these metrics is a computationally intensive task, motivating the need for dimensionality reduction in advance [313]. For this purpose, we use the t-distributed stochastic neighbor embedding (t-SNE) [314]. As such, clustering-based analyses were conducted on the reduced-dimensionality samples, and the results are summarized in Fig. 4.16. As described in this figure, the highest and lowest clustering performances were achieved for the FECGSYNDB and NIFECGC datasets with average (± standard deviation) purity indices of 0.9992 (± 0.0005) and 0.9412 (± 0.0279) respectively, whereas a purity index of 0.9846 (± 0.0072) was achieved for ADFECGDB, implying high concentration of the dominant samples (either maternal or fetal) in their respective cluster. Similarly, the Jaccard values of 0.9984 (± 0.0194), 0.0.9779 (± 0.0453), and 0.9352 (± 0.0388) respectively suggest high similarities between FECGSYNDB,

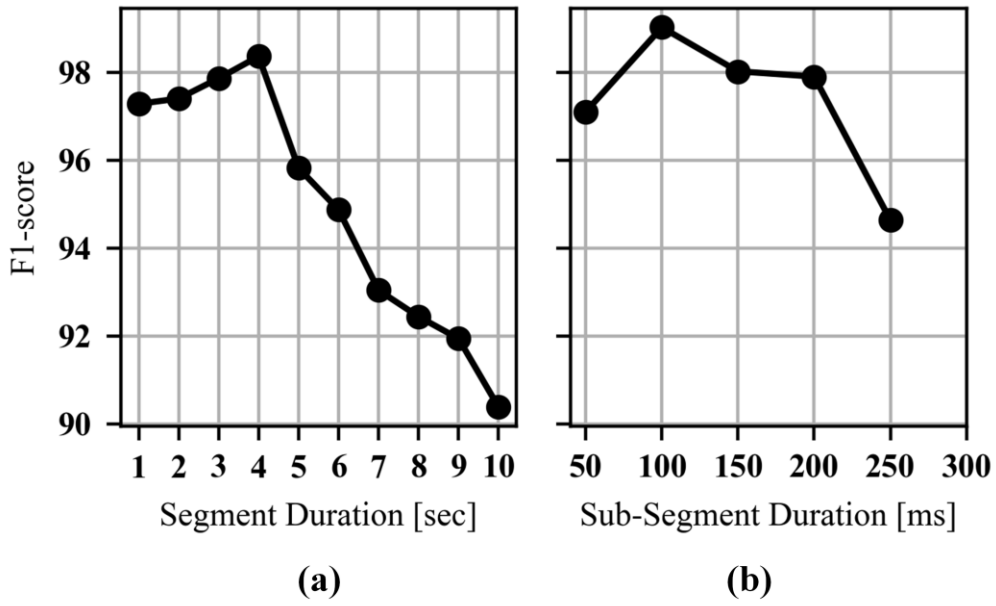


Fig. 4.17 Impacts of (a) segment duration and (b) sub-segment duration on the F1-score obtained for the detection of fetal QRS complexes in FECGSYNDB.

ADFECGDB, and NIFECGC datasets and their respective ground-truth labels. Moreover, the Davies-Bouldin (DB) values signify the separability of the clusters for FECGSYND, ADFECGDB, and NIFECGC by 0.7156 (± 0.0721), 0.7318 (± 0.0611), and 0.7813 (± 0.0851) respectively. It is to be noted that the lower the Davies-Bouldin metric, the higher the separability. According to the achieved Davies-Bouldin values, the intra-cluster distances are smaller than their respective inter-cluster distances, suggesting separable fetal and maternal masking maps. The results of this study were published in [299].

#### 4.5.3.5 Ablation Study

A common practice to better generalize a neural network to unseen data (real datasets) is by an ablation study on the constituent parts of the training procedure. In this study, we investigate the impacts of the denoising module (DN), the number of dual-path LSTM modules (DP), the sparse feature extraction mechanism (SF), the data augmentation (DA), and the lengths of segments and sub-segments on the performance of the DPSS network. The ablation study is conducted for each fold of FECGSYNDB and averaged over all 10 subjects to maximize the network F1-score.

Comparing architectures 1 and 2 in Table 4.6 illustrates the impact of the denoising network on the F1-score (61.76% vs. 84.21%), suggesting a 22.45% improvement as a result of incorporating the denoising network. According to architectures 1-5, including 3 dual-path LSTM (DP) modules leads to the superior performance among all similar architectures with 1, 2, 3, and 4 DP module(s). An F1-score of 85.13% demonstrates the superiority over the runner-up architecture with 1 DP module and F1-score of 84.21%. The

weaker performance in architecture 5 compared to Baseline (1 DP) + DN can be related to the higher complexity of the model, which causes the network to overfit the data.

A comparison conducted among architectures 6-11 implies that the model with 6-filter sparse feature extraction, which is referred to as SF6, offers the highest F1-score (94.68%). It should be noted that architecture Baseline (3 DP + SF6) + DN leverages filters with sizes of 1, 3, 5, 7, 9, and 13. Intuitively, employing filters with different sizes allows for extracting both fine and coarse features which help the neural network learn the

Table 4.6 Ablation Study Based on the Performance for Fetal ECG Extraction from FECGSYNDB

| Architecture | F1-score (%) |
|---|---|
| **1. Baseline (1 DP)** | **61.67 (± 1.52)** |
| **2. Baseline (1 DP) + DN** | **84.21 (± 1.06)** |
| 3. Baseline (2 DP) + DN | 84.10 (± 1.17) |
| **4. Baseline (3 DP) + DN** | **85.13 (± 0.96)** |
| 5. Baseline (4 DP) + DN | 82.10 (± 0.82) |
| 6. Baseline (3 DP + SF3) + DN | 88.73 (± 0.45) |
| 7. Baseline (3 DP + SF4) + DN | 90.16 (± 0.48) |
| 8. Baseline (3 DP + SF5) + DN | 93.07 (± 0.68) |
| **9. Baseline (3 DP + SF6) + DN** | **94.68 (± 0.39)** |
| 10. Baseline (3 DP + SF7) + DN | 92.43 (± 0.93) |
| 11. Baseline (3 DP + SF8) + DN | 91.77 (± 0.81) |
| **12. Baseline (3 DP + SF6) + DN + DA** | **99.03 (± 0.39)** |

DP: Dual-path LSTM, DN: denoising, SF: sparse feature extraction, and DA: data augmentation.

distribution mode deeply. Nevertheless, increasing the number of filters results in more complex architectures that might overfit the distribution of the data. Two examples of overfitting are observed in Table 4.6 for architecture 10 with SF7 and architecture 11 with SF8 with F1-scores of 92.43% and 91.77%, respectively.

A comparison between architectures 9 and 12 signifies the importance of data augmentation (DA). For this part, the dataset is augmented to achieve a twice larger dataset than the original size (3,432,960 vs. 1,716,480 segments). As reported in Table 4.6, the Baseline (3 DP + SF6) + DN + DA outperforms the model without data augmentation (Baseline (3 DP + SF6) + DN) by 4.35% F1-score.

Finally, the impacts of the segment length ($T$) and sub-segment length ($L$) are investigated in Figs 4.17 (a) and (b) respectively. According to Fig. 4.17 (a), the model is trained for segment lengths within the range of 1-10 seconds, where the highest average F1-score of 98.37 % is achieved for a segment length of 4 seconds. As can be seen in this figure, the average F1-score decreases from 98.37% to 90.39% when moving from 4 seconds to 10 seconds. Fig. 4.17 (b) depicts the average F1-score in terms of sub-segment duration, where 100 ms suggests a higher F1-score compared to other durations (99.03% vs. 97.11%, 98.03%, 97.91%, and 94.65% for 50 ms, 150 ms, 200 ms, and 250 ms respectively).

#### 4.5.3.6 Comparative Study on Fetal QRS Detection and FHR Estimation

In Table 4.7, DPSS is compared with state-of-the-art single-channel fetal ECG extraction methods. As summarized in Table 4.7, an F1-score of 95.3% on the NIFECGC dataset suggests the superiority of DPSS over the previous works except [297], [260],

[285], and [286]. Authors in [297] and [260] have used all 4 abdominal channels, whereas DPSS requires only a single channel to reconstruct FECG signal. Performance achieved in [285] and [286] represent 17 and 27 subjects, respectively, whereas our work is evaluated on all 69 subjects with correct signal annotations. In terms of FHR precision, DPSS suggests 88.61% and 83.86% for ADFECGDB and NIFECGC, respectively, outperforming all previous works which have reported on FHR precision values. In addition, DPSS shows promising results on the ADFECGDB dataset with an F1-score of 97.7% on 22 signals, which is a superior estimation to those conventional methods except [290], [286], [285], and [272] with 99.7%, 98.04%, 99.27%, and 99.44% respectively. Although DPSS provides slightly weaker performance, it should be noted that we have used the extended version of ADFECGDB (labor and pregnancy) in contrast to the literature despite its low computational demands, which will be indicated in the following section. literature. In general, DPSS suggests quite an acceptable performance in

comparison to literature despite its low computational demands, which will be indicated in the following section.

### 4.5.4 Discussion

As demonstrated in earlier sections, DPSS is capable of fetal ECG extraction from abdominal recordings. A distinctive characteristic of DPSS is its robustness against coincided fetal and maternal QRS peaks, which often inhibits conventional methods from

Table 4.7 Comparative Evaluation for the Detection of Fetal QRS Complexes and FHR Estimation vs. State-of-the-Art Methods

| Method | Dataset | Channel(s) | Recordings | Year | FHR Precision (%) | Sen (%) | PPV (%) | Acc (%) | F1 (%) |
|---|---|---|---|---|---|---|---|---|---|
| TS (Cerutti et al.) | NIFECG | 1 | 14 | 1986 | 57.9 | 91.2 | 90.5 | NA | 90.8 |
| | Private | | | | 68.1 | 86.8 | 85.3 | | 86 |
| TS + PCA (Kanjilal et al.) | NIFECG | 1 | 14 | 1997 | 68.7 | 94.7 | 96 | NA | 95.4 |
| | Private | | | | 73.9 | 89.9 | 88.8 | | 89.3 |
| LMS (Behar et al.) | NIFECG | 1 | 14 | 2014 | 69.3 | 95.8 | 95 | NA | 95.4 |
| | Private | | | | 73.7 | 89.3 | 86.5 | | 87.9 |
| RLS (Behar et al.) | NIFECG | 1 | 14 | 2014 | 70.6 | 96.2 | 95.6 | NA | 95.9 |
| | Private | | | | 74.5 | 89.7 | 86.8 | | 88.2 |
| ESN (Behar et al.) | NIFECG | 1 | 14 | 2014 | 72.2 | 96.8 | 97.2 | NA | 97 |
| | Private | | | | 78.7 | 91.4 | 88.9 | | 90.2 |
| Behar et al. | NIFECGC | 4 | 69 | 2014 | NA | 95.9 | 96 | NA | 96 |
| Varanin et al. | NIFECGC | 4 | 69 | 2014 | NA | 99.1 | 98.9 | NA | 98.99 |
| SVD + SW (Zhang et al.) | ADFECGDB | 1 | 2 | 2017 | NA | 99.36 | 99.52 | 98.89 | 99.44 |
| | Private | | 1 | | | 98.31 | 98.86 | 97.21 | 98.58 |
| STFT + NMF (Gurve et al.) | ADFECGDB | 1 | 5 | 2019 | NA | 95.3 | 94.6 | NA | 94.8 |
| | NIFECGC | | 60 | | | NA | NA | | 83.8 |
| Mohebbian et al. | NIFECG | 1 | 14 | 2020 | NA | 95.3 | 97.1 | NA | 96.1 |
| TF + Shannon (Krupa et al.) | ADFECGDB | 1 | 13 | 2022 | NA | 99.37 | 99.38 | 98.55 | 99.27 |
| | NIFECGC | | 26 | | | 98.61 | 98.72 | 97.37 | 98.67 |
| Clustering (Zhang et al.) | ADFECGDB | 1 | 11 | 2018 | NA | 97.5 | 98.58 | NA | 98.04 |
| | NIFECGC | | 64 | | | 97.51 | 98.63 | | 98.07 |
| RCED-NET (Zhong et al.) | ADFECGDB | 1 | 9 | 2019 | NA | 96.06 | 92.25 | NA | 94.1 |
| | NIFECGC | | 40 | | | 92.6 | 94.68 | | 93.62 |
| CycleGAN (Mohebbian et al.) | ADFECGDB | 1 | 5 | 2021 | NA | 99.4 | 99.6 | NA | 99.7 |
| | NIFECG | | 14 | | | 96.8 | 97.2 | | 97.9 |
| | NIFECGC | | 69 | | | 93.9 | 93.9 | | 94.7 |
| STFT + GAN (Zhong et al.) | ADFECGDB | 1 | 5 | 2021 | NA | 90.32 | 87.79 | NA | 90.05 |
| | NIFECGC | | 80 | | | 92.37 | 93.69 | | 93.02 |
| DPSS (Proposed) | **ADFECGDB(P)** | **1** | **10** | **2022** | **87.88** | **96.5** | **97.98** | **94.67** | **97.24** |
| | **ADFECGDB(L)** | | **12** | | **89.22** | **98** | **98.18** | **96.25** | **98.08** |
| | **ADFECGDB(P)** | | **22** | | **88.61** | **97.3** | **98.09** | **95.53** | **97.7** |
| | **NIFECGC** | | **69** | | **83.86** | **94.2** | **96.5** | **91.34** | **95.3** |

TS: template subtraction, PCA: principal component analysis, LMS: least mean-square, RLS: recursive mean-square, ESN: echo state network, SVD: singular value decomposition, SW: smooth window, STFT: short-time Fourier transform, NMF: non-negative matrix factorization, TF: time-frequency, RCED-NET: residual convolutional encoder-decoder network, GAN: generative adversarial network, and DPSS: dual-path source separation.

detecting fetal peaks. This is explained by a GAN-based denoiser which cancels noise components by learning the distribution of the abdominal signals. Furthermore, the results of this study demonstrate that masking functions, directly extracted from the abdominal signal, allow for accurately separating fetal components in time domain. Fig. 4.18 provides a 30-second example of recording a016 from the NIFECGC dataset. As can be observed in Fig. 4.18 (c), the green rectangle signifies an example of fetal and maternal QRS peaks, where masking functions have successfully separated the sources. Additionally, a significant finding of this study is fetal QRS amplification by denoiser. A comparison between fetal QRS peaks in Figs 4.18 (a) and (b) indicates that fetal QRS peaks are amplified in terms of amplitude as a result of the denoising stage.

Despite the promising results achieved on real datasets, the performance of DPSS is limited when the GAN-based denoiser enhances artifacts that resemble fetal QRS peaks. As demonstrated in the red panels of Fig. 4.18, DPSS might detect noise components as fetal QRS peaks, leading to false positive cases. Fetal QRS peaks might also be missed by the denoiser as a result of high-energy noise, an example of which is illustrated in the brown panel of Fig. 4.18 (a). According to its counterpart in Fig. 4.18 (b), the whole segment is smoothed, eliminating fetal QRS waveform and increasing false negatives. Another finding about the extracted fetal ECG signals is the oscillatory behavior in between consecutive QRS peaks which contaminates fetal QRS P and T peaks. According to our observation, this phenomenon further degrades the quality of fetal ECG signals when it is accompanied by the remnant of maternal QRS peaks.

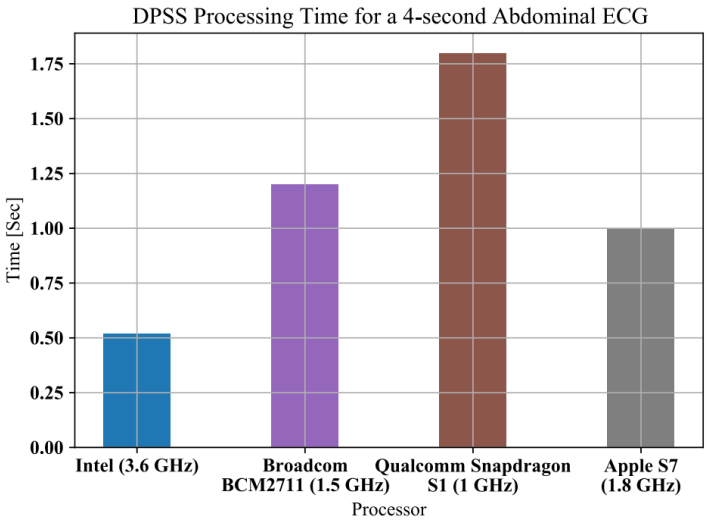


Fig. 4.19 Processing time of a 4-second abdominal ECG segment on different processors.

To have a real-time algorithm, the processing time of a signal segment should be less than the segment length. In order to investigate the real-time practicality of DPSS, we transferred the trained DPSS network to an Intel processing unit of a computer with a 3.6 GHz clock speed. It was observed that the processing unit for a 4-second abdominal signal was 0.52 s, as illustrated in Fig. 4.19. As demonstrated in this figure, the processing time was also calculated for off-the-shelf processors, such as Broadcom BCM2711 (in Raspberry Pi), Qualcomm Snapdragon S1 (in smartphones), and Apple S7 (in Apple Watch), the processing times of which are 1.2 s, 1.8 s, and 1 s respectively. These intervals, which are less than 4 seconds, indicate that DPSS holds low computational complexity which makes it a suitable candidate for real-time fetal monitoring systems.

# 5 Conclusion

This dissertation has investigated the use of wearable sensors for cardiovascular monitoring and fetal heart rate extraction. The results demonstrate that wearable sensors are a reliable technology for the continuous monitoring of cardiovascular function and fetal heart rate, providing an efficient and non-invasive method for clinical settings as well as home-based use.

The second chapter described the development of a new reference-free framework that uses SCG and GCG morphological characteristics and HRV parameters to detect aortic stenosis. The framework included optimized filter design and the preparation of two groups of datasets at the subject and chunk levels. Additionally, new parameters called MA and P-included intervals were introduced, which demonstrated higher consistency with AS compared to other top-ranked features. Although time-domain and frequency-domain HRV parameters were extracted, there was only a low correlation between HRV parameters and AS. In contrast, ML models trained on GCG timing intervals showed perfect discrimination between AS and non-AS groups. XGBoost is the most accurate ML model for both subject-level and chunk-level analyses, with F1-scores of 100.00% and 96.49%, respectively. The optimized filter design was shown to be effective at both levels, leading to better performance compared to previous literature. Lastly, the proposed framework was also able to classify the severity level of AS with 93.80% accuracy and F1-score. These findings demonstrated the robustness and potential of the proposed framework for AS detection and severity classification.

The second chapter also presented a novel framework for the detection of coexisting valvular heart diseases (VHDs) using a high-precision sensitive accelerometer contact microphone (ACM). This sensor enabled the capturing of minute heart murmurs associated with valvular abnormalities to differentiate between different VHD conditions and healthy individuals. In order to detect abnormal patterns in ACM recordings, we developed an image-to-sequence translation neural network that took advantage of both convolutional neural networks (CNNs) and transformers, i.e., "convolution-meets-transformer (CMT)". This architecture enables the modeling of local and global dependencies in Mel-frequency cepstral coefficients (MFCCs) of the input signal, and translated them to a 5-digit binary sequence in the output. Each digit corresponded to either aortic stenosis, aortic regurgitation, mitral regurgitation, pulmonary regurgitation, or tricuspid regurgitation.

The efficacy of the proposed framework was evaluated on the data of 110 subjects consisting of 58 VHD patients and 52 healthy individuals. Statistical analyses on the test set suggested average (± std.) sensitivity, specificity, accuracy, positive predictive value, and F1 score of 93.28% (± 4.88), 98.07% (± 1.97), 96.87% (± 2.08), 92.97% (± 7.13), and 92.94% (± 4.69) respectively. Comparative analyses indicated that the ACM MFCC + CMT network offered superior performance compared to state-of-the-art methods for VHD detection, and also for the identification of coexisting diseases which was not possible by conventional methods.

Additionally, a novel framework based on the proximal recordings of heart vibrations using a high-precision accelerometer contact microphone (ACM) for the detection of peripheral artery disease (PAD) and the classification of its severity level was presented.

A deep neural network coined multi-stream-powered vision transformer (MSPViT) was introduced to differentiate PAD patients from healthy subjects using blood pressure backflow patterns in ACM recordings corresponding to arterial obstructions. The proposed architecture leveraged Mel frequency cepstral coefficients (MFCCs) and high-level-of-abstraction coefficients (HLACs) generated by a pre-trained ResNet50 network for PAD detection and severity classification. The performance of the proposed framework was evaluated on the data of 74 PAD patients and 21 healthy subjects (a total of 95 subjects) for PAD detection as well as severity classification. Sensitivity, specificity, accuracy, positive predictive value, and F1 scores of 99.45% (± 0.47), 98.21% (± 2.93), 99.07% (± 0.89), 99.30% (± 1.25), and 99.37% (± 0.63) were respectively reported for PAD detection using MSPViT (X2) whose input was provided by layer 22 of the fine-tuned ResNet50 network. An AUC of 0.99 also confirmed the discrimination power of MSPViT (X2). Additionally, the network was assessed for the classification of severity levels where average sensitivity, specificity, accuracy, positive predictive value, and F1 scores of 96.66% (± 4.08), 97.34% (± 2.19), 97.65% (± 1.67), 96.24% (± 2.51), and 96.29% (± 3.01) were achieved.

The chapter ends with introducing the concept of continuous blood pressure estimation based on strain plethysmography (SPG) recordings at the fingertip. The experimental setup took advantage of a custom-built MEMS strain sensor to collect pulse information from the digital artery at the fingertip. A comprehensive feature set including timing intervals, amplitude, phase, statistical, and demographic information were extracted. Furthermore, geometric characteristics of the triangles formed by the fiducial points of the signal and its

derivatives along with the use of the first and second-order integrals were introduced for feature extraction. An XGBoost regressor model was employed to relate the extracted feature space to the systolic blood pressure (SBP), diastolic blood pressure (DBP), and mean arterial pressure (MAP) values. The performance of the proposed framework was evaluated on the data of 32 subjects using 5-fold cross-validation (5-CV) and subject-specific approaches. The fingertip SPG-based method demonstrated reliability for BP monitoring, with mean absolute errors (MAEs) of 3.88 mmHg and 5.45 mmHg achieved for DBP and SBP estimation in 5-CV, and 3.70 mmHg and 5.03 mmHg in subject-specific validation, respectively. Three geometric and 26 integral-related features were identified as top contributors for BP estimation. The method showed robustness against blood pressure level and body mass index (BMI), with MAE ranges of [1.28, 4.28] mmHg and [2.64, 7.52] mmHg, respectively.

The third chapter of the dissertation described a new approach for detecting left ventricular abnormalities. The method involved using an accelerometer contact microphone (ACM) combined with a single-lead electrocardiogram (ECG) sensor to estimate the pre-ejection period (PEP) and the left ventricular ejection time (LVET), which were then used to classify between healthy subjects and those with LVEF abnormalities. The method was tested on a group of 8 healthy individuals and 7 CVD patients with abnormal LVEF values, and showed a promising accuracy of 84.47% for use in wearable settings.

Additionally, this chapter presented a novel framework for classifying the severity level of aortic stenosis based on changes in the pattern of angular chest movements induced

by cardiac activity. A range of timing intervals between fiducial points on the GCG signal and time-domain HRV parameters were used to create a large feature space, and machine learning techniques were used to classify the severity of stenosis. The LightGBM classifier outperformed others with an F1-score of 94.29%, and the top features are ranked using the SHAP method, where IVCT and IVRT were among the top frequencies.

Furthermore, this chapter reported on fingertip strain plethysmography (SPG) for cardiovascular monitoring. The developed method took advantage of a small, sensitive MEMS strain sensor to capture pulsations on the fingertip. These pulsations were related to structural changes in blood vessels in each heartbeat. The proposed sensing methodology was assessed on 10 healthy subjects for the estimation of hemodynamic parameters, including heart rate (HR), heart rate variability (HRV), and respiratory rate (RR). It was demonstrated that SPG could provide reliable estimations of HR, HRV, and RR with accuracies of 99.88%, 96.43%, and 92.64% respectively. Additionally, this chapter compared the estimation of HRV parameters from digital (fingertip) and radial (wrist) arteries. A highly-sensitive micro-electromechanical systems (MEMS) strain plethysmography sensor (SPG) was utilized to capture pulse information from the fingertip and wrist of 15 subjects. The recordings were used to estimate heart rate variability (HRV) parameters, such as meanNN, meanHR, SDNN, RMSSD, PNN50, and LF/HF. It was demonstrated that SPG-based methods could provide information that can be acquired with PPG sensors as well, while eliminating optical limitations. It was also demonstrated that meanNN, SDNN, RMSSD, PNN50, and LF/HF can be estimated from SPG recordings with mean differences of 0.0 ms, 0.00 BPM, 1 ms, 1.6 ms, 0.5747%, and 0.0332,

respectively, for digital artery, and mean differences of 0.1 ms, 0.00 BPM, 1 ms, 3.5 ms, 5.11%, and -0.0151, respectively, for radial artery. This set of evidence suggested that fingertip is a more reliable arterial site for the estimation of HRV parameters.

The final section of the chapter was dedicated to the estimation of pulse pressure from central recordings of accelerometer contact microphone (ACM). This study explored the relation between pulse pressure and heart sounds acquired by a sensitive accelerometer contact microphone (ACM). It is demonstrated that blood pulse pressure can be predicted using the features derived from the detail coefficients of the discrete wavelet transform with accuracies of 77.14% and 95.68% for leave-subject-out and 10-fold cross-validations, respectively. Furthermore, it was proven that while extracted features from high-frequency bands such as 1-2 kHz and 512-1,204 Hz are the most consistent with pulse pressure values, medium frequency bands such as 64-128 Hz and 32-64 Hz were also contributive to the estimated values.

The fourth chapter addressed the fusion of FECG and inertial sensors to monitor FHR. The proposed framework benefited from an EMD-based signal refinement, which enhanced fetal cardiac activity components in both ECG and SCG/GCG readings. Furthermore, two sensor fusion methods were designed to improve the accuracy of FHR estimation. A PPA of 87.00% demonstrated the efficiency of our approach for FHR extraction.

This chapter also reported on the development of a signal-to-signal translation technique for the cancellation of fetal movement in abdominal ECG recordings. The proposed method performed based on a one-to-one mapping between two domains

corresponding to noisy abdominal ECG signals and their clean counterparts. For this purpose, a generative adversarial network (GAN) was leveraged to ensure discovering cross-domain relations between the two domains. Two generators were trained, one for mapping raw abdominal ECG signals to their clean counterparts, and the other for translating clean abdominal ECG signals to their noisy versions. The model achieved average MAE, RMSD, and PCC values of 0.099, 0.124, and 99.12% respectively on the PhysioNet fetal ECG synthetic database. Furthermore, the robustness of the method was evaluated in terms of input SNR values, where the RMSD value varies within the range of (0.047, 0.352). This indicated high correlations between ground-truth clean signals and their denoised counterparts.

The last section of the chapter introduced DPSS, a novel method for fetal ECG (FECG) extraction from single-channel abdominal ECG (AECG) recordings. Benefiting from the dual-path long short-term memory (DP-LSTM) mechanism, the DPSS framework estimated FECG and MECG signals by masking raw AECG recordings. We evaluated the performance of DPSS on the detection of fetal QRS complexes, and demonstrate the F1-scores of 99.03%, 95.3%, and 97.7% on FECGSYNDB with 10 simulated pregnancies, NIFECGC with 69 actual pregnancies, and ADFECGDB with 22 actual labor and pregnancies, respectively. The separability of maternal and fetal masking maps was shown by average purity index of 0.975, Jaccard index of 0.9705, and Davies-Bouldin index of 0.7429 respectively. According to the ablation study on the network configuration, the architecture with the denoising module, three dual-path LSTM's, and 6 feature extraction filters, along with data augmentation was found to offer the best performance (F1-score of

99.03%). Additionally, this study demonstrated that the optimum time durations of signal segments and sub-segments for DPSS are 4 seconds and 100 ms respectively.

Overall, this dissertation has contributed to the growing body of research on wearable sensors for healthcare monitoring. The results provide important insights into the potential benefits and challenges associated with the use of wearable sensors for cardiovascular monitoring and fetal heart rate extraction, and suggest that wearable sensors have the potential to revolutionize the way we monitor and manage cardiovascular and fetal health.

## Bibliography

[1] Plana JC (2017) Valvular Heart Disease. *Anticancer Treatments and Cardiotoxicity: Mechanisms, Diagnostic and Therapeutic Interventions*, https://doi.org/10.1016/B978-0-12-802509-3.00018-2

[2] Members WC, Otto CM, Nishimura RA, Bonow RO, Carabello BA, Erwin III JP, Gentile F, Jneid H, Krieger E V, Mack M (2021) 2020 ACC/AHA guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Journal of the American College of Cardiology*, 77(4):e25–e197.

[3] Veronesi C, Beccagutti G, Corbo M, Blini V, Degli Esposti L (2015) Cost Of Illness In Aortic Stenosis Patients. *Value in Health*, 18(7):A386.

[4] Mitchell C, Rahko PS, Blauwet LA, Canaday B, Finstuen JA, Foster MC, Horton K, Ogunyankin KO, Palma RA, Velazquez EJ (2019) Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography. *Journal of the American Society of Echocardiography*, 32(1):1–64.

[5] Bradley SM, Foag K, Monteagudo K, Rush P, Strauss CE, Gössl M, Sorajja P (2019) Use of routinely captured echocardiographic data in the diagnosis of severe aortic stenosis. *Heart*, 105(2):112–116.

[6] Gardezi SKM, Myerson SG, Chambers J, Coffey S, d'Arcy J, Hobbs FDR, Holt J, Kennedy A, Loudon M, Prendergast A (2018) Cardiac auscultation poorly predicts the presence of valvular heart disease in asymptomatic primary care patients. *Heart*, 104(22):1832–1835.

[7] Cleveland Clinic (2022) Echocardiogram. https://my.clevelandclinic.org/health/diagnostics/16947-echocardiogram

[8] Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Cheng S, Delling FN (2021) Heart disease and stroke statistics—2021 update: a report from the American Heart Association. *Circulation*, 143(8):e254–e743.

[9] Criqui MH, Aboyans V (2015) Epidemiology of peripheral artery disease. *Circulation research*, 116(9):1509–1526.

[10] McDermott MM, Fried L, Simonsick E, Ling S, Guralnik JM (2000) Asymptomatic peripheral

arterial disease is independently associated with impaired lower extremity functioning: the women's health and aging study. *Circulation*, 101(9):1007–1012.

[11] Selvin E, Erlinger TP (2004) Prevalence of and risk factors for peripheral arterial disease in the United States: Results from the National Health and Nutrition Examination Survey, 1999-2000. *Circulation*, 110(6):738–743. https://doi.org/10.1161/01.CIR.0000137913.26087.F0

[12] Kullo IJ, Rooke TW (2016) Peripheral artery disease. *New England Journal of Medicine*, 374(9):861–871.

[13] Carter SA (1968) Indirect systolic pressures and pulse waves in arterial occlusive disease of the lower extremities. *Circulation*, 37(4):624–637.

[14] Rac-Albu M, Iliuta L, Guberna SM, Sinescu C (2014) The role of ankle-brachial index for predicting peripheral arterial disease. *Maedica*, 9(3):295–302. http://www.ncbi.nlm.nih.gov/pubmed/25705296%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4306002

[15] Lijmer JG, Hunink MGM, Dungen JJAM van den, Loonstra J, Smit AJ (1996) ROC analysis of noninvasive tests for peripheral arterial disease. *Ultrasound in medicine & biology*, 22(4):391–398.

[16] Taggart NW, Cabalka AK (2016) Cardiac catheterization and angiography. *Moss and Adams' Heart Disease in Infants, Children, and Adolescents: Including the Fetus and Young Adult: Ninth Edition*, 1

[17] Schernthaner R, Stadler A, Lomoschitz F, Weber M, Fleischmann D, Lammer J, Loewe C (2008) Multidetector CT angiography in the assessment of peripheral arterial occlusive disease: accuracy in detecting the severity, number, and length of stenoses. *European radiology*, 18(4):665–671.

[18] Laswed T, Rizzo E, Guntern D, Doenz F, Denys A, Schnyder P, Qanadli SD (2008) Assessment of occlusive arterial disease of abdominal aorta and lower extremities arteries: Value of multidetector CT angiography using an adaptive acquisition method. *European Radiology*, 18(2):263–272. https://doi.org/10.1007/s00330-007-0749-0

[19] Gradinscak DJ, Young N, Jones Y, O'Neil D, Sindhusake D (2004) Risks of outpatient angiography and interventional procedures: a prospective study. *American Journal of*

*Roentgenology*, 183(2):377–381.

[20] Forghani N, Maghooli K, Dabanloo NJ, Farahani AV, Forouzanfar M (2022) DeepPAD: Detection of Peripheral Arterial Disease Using Deep Learning. *IEEE Sensors Journal*,

[21] Forghani N, Maghooli K, Jafarnia Dabanloo N, Vasheghani Farahani A, Forouzanfar M (2021) Intelligent Oscillometric System for Automatic Detection of Peripheral Arterial Disease. *IEEE Journal of Biomedical and Health Informatics*, 25(8):3209–3218. https://doi.org/10.1109/JBHI.2021.3065379

[22] Allen J (2007) Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1.

[23] Allen J, Overbeck K, Nath AF, Murray A, Stansby G (2008) A prospective comparison of bilateral photoplethysmography versus the ankle-brachial pressure index for detecting and quantifying lower limb peripheral arterial disease. *Journal of vascular surgery*, 47(4):794–802.

[24] Peltokangas M, Telembeci AA, Verho J, Mattila VM, Romsi P, Vehkaoja A, Lekkala J, Oksala N (2017) Parameters extracted from arterial pulse waves as markers of atherosclerotic changes: performance and repeatability. *IEEE journal of biomedical and health informatics*, 22(3):750–757.

[25] Bentham M, Stansby G, Allen J (2018) Innovative multi-site photoplethysmography analysis for quantifying pulse amplitude and timing variability characteristics in peripheral arterial disease. *Diseases*, 6(3):81.

[26] Allen J, Liu H, Iqbal S, Zheng D, Stansby G (2021) Deep learning-based photoplethysmography classification for peripheral arterial disease detection: a proof-of-concept study. *Physiological Measurement*, 42(5):54002.

[27] Wu J-X, Li C-M, Ho Y-R, Wu M-J, Huang P-T, Lin C-H (2015) Bilateral photoplethysmography analysis for peripheral arterial stenosis screening with a fractional-order integrator and info-gap decision-making. *IEEE Sensors Journal*, 16(8):2691–2700.

[28] Yousef Q, Reaz MBI, Ali MAM (2012) The analysis of PPG morphology: investigating the effects of aging on arterial compliance. *Measurement Science Review*, 12(6):266.

[29] Rubins U, Grabovskis A, Grube J, Kukulis I (2008) Photoplethysmography analysis of artery properties in patients with cardiovascular diseases. *14th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics*, :319–322.

[30] Son M, Miller ES (2017) Predicting preterm birth: cervical length and fetal fibronectin. *Seminars in perinatology*, 41(8):445–451.

[31] Hezelgrave NL, Shennan AH (2016) Quantitative fetal fibronectin to predict spontaneous preterm birth: a review. *Women's Health*, 12(1):121–128.

[32] Brown R, Wijekoon JHB, Fernando A, Johnstone ED, Heazell AEP (2014) Continuous objective recording of fetal heart rate and fetal movements could reliably identify fetal compromise, which could reduce stillbirth rates by facilitating timely management. *Medical hypotheses*, 83(3):410–417.

[33] Goldenberg RL, Kirby R, Culhane JF (2004) Stillbirth: a review. *The journal of maternal-fetal & neonatal medicine*, 16(2):79–94.

[34] United Nations IGME agencies (2020) A neglected tragedy : the global burden of stillbirths. https://www.unicef.org/reports/neglected-tragedy-global-burden-of-stillbirths-2020

[35] KALACHE KD, DÜCKELMANN AM (2012) Doppler in Obstetrics. *Clinical Obstetrics & Gynecology*, 55(1):288–295. https://doi.org/10.1097/grf.0b013e3182488156

[36] Avoid Fetal "Keepsake" Images, Heartbeat Monitors. https://www.fda.gov/ForConsumers /ConsumerUpdates/ucm095508.htm

[37] Vullings R, Peters CHL, Mischi M, Oei SG, Bergmans JWM (2008) Fetal movement quantification by fetal vectorcardiography: a preliminary study. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, :1056–1059.

[38] Graatsma M (2010) Monitoring of fetal heart rate and uterine activity. *IMJ. Illinois medical journal*, 149(1)http://www.ncbi.nlm.nih.gov/pubmed/2544

[39] Doppler B (2021) Sonoline B - Baby Doppler - Fetal Monitor (Pink). https://www.babydoppler.com/sonoline-b-fetal-doppler-pink.html

[40] Majumder S, Mondal T, Deen MJ (2017) Wearable sensors for remote health monitoring. *Sensors*, 17(1):130.

[41] Behar J, Oster J, Clifford GD (2014) Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data. *Physiological measurement*, 35(8):1569.

[42] Kovács F, Horváth C, Balogh ÁT, Hosszú G (2011) Fetal phonocardiography—past and future

possibilities. *Computer methods and programs in biomedicine*, 104(1):19–25.

[43] Yang C, Antoine C, Young BK, Tavassolian N (2019) A Pilot Study on Fetal Heart Rate Extraction from Wearable Abdominal Inertial Sensors. *IEEE Sensors Journal*, 19(22):10773–10781. https://doi.org/10.1109/JSEN.2019.2930886

[44] Shokouhmand A, Antoine C, Young BK, Tavassolian N (2021) Multi-modal Framework for Fetal Heart Rate Estimation: Fusion of Low-SNR ECG and Inertial Sensors. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, :1–4.

[45] Kahankova R, Martinek R, Jaros R, Behbehani K, Matonia A, Jezewski M, Behar JA (2019) A review of signal processing techniques for non-invasive fetal electrocardiography. *IEEE reviews in biomedical engineering*, 13:51–73.

[46] Behar J, Andreotti F, Zaunseder S, Oster J, Clifford GD (2016) A practical guide to non-invasive foetal electrocardiogram extraction and analysis. *Physiological Measurement*, 37(5):R1–R35. https://doi.org/10.1088/0967-3334/37/5/R1

[47] Martinek R, Kahankova R, Nazeran H, Konecny J, Jezewski J, Janku P, Bilik P, Zidek J, Nedoma J, Fajkus M (2017) Non-invasive fetal monitoring: A maternal surface ECG electrode placement-based novel approach for optimization of adaptive filter control parameters using the LMS and RLS algorithms. *Sensors*, 17(5):1154.

[48] Wang Y, Fu Y, He Z (2018) Fetal Electrocardiogram Extraction Based on Fast ICA and Wavelet Denoising. *Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2018*, 2:466–469. https://doi.org/10.1109/IMCEC.2018.8469501

[49] Liu G, Luan Y (2015) An adaptive integrated algorithm for noninvasive fetal ECG separation and noise reduction based on ICA-EEMD-WS. *Medical & biological engineering & computing*, 53(11):1113–1127.

[50] Sameni R, Clifford GD, Jutten C, Shamsollahi MB (2007) Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals. *Eurasip Journal on Advances in Signal Processing*, 2007:1–14. https://doi.org/10.1155/2007/43407

[51] Comon P, Jutten C (2010) Handbook of Blind Source Separation: Independent component analysis and applications.

[52] Schaefer BM, Otto CM (2009) Aortic Stenosis. *Cardiovascular Imaging: A Handbook for Clinical Practice*, 373(9667):26–35. https://doi.org/10.1002/9781444312898.ch3

[53] Grimard BH, Larson JM (2008) Aortic stenosis: diagnosis and treatment. *American family physician*, 78(6):717–724.

[54] Makkar RR, Fontana GP, Jilaihawi H, Kapadia S, Pichard AD, Douglas PS, Thourani VH, Babaliaros VC, Webb JG, Herrmann HC (2012) Transcatheter aortic-valve replacement for inoperable severe aortic stenosis. *New England Journal of Medicine*, 366(18):1696–1704.

[55] Saji M, Higuchi R, Izumi Y, Takamisawa I, Iguchi N, Nanasato M, Shimizu J, Shimokawa T, Takanashi S, Takayama M, Isobe M (2021) Transcatheter aortic valve replacement in patients with extremely severe aortic stenosis. *International Journal of Cardiology*, 329(18):162–166. https://doi.org/10.1016/j.ijcard.2020.12.080

[56] Svensson LG (2008) Aortic valve stenosis and regurgitation: an overview of management. *Journal of Cardiovascular Surgery*, 49(2):297.

[57] Rosenhek R, Binder T, Porenta G, Lang I, Christ G, Schemper M, Maurer G, Baumgartner H (2000) Predictors of Outcome in Severe, Asymptomatic Aortic Stenosis. *New England Journal of Medicine*, 343(9):611–617. https://doi.org/10.1056/nejm200008313430903

[58] Vaquette B, Corbineau H, Laurent M, Lelong B, Langanay T, Place C De, Froger-Bompas C, Leclercq C, Daubert C, Leguerrier A (2005) Valve replacement in patients with critical aortic stenosis and depressed left ventricular function: predictors of operative risk, left ventricular function recovery, and long term outcome. *Heart*, 91(10):1324–1329.

[59] Otto CM (2006) Valvular aortic stenosis: disease severity and timing of intervention. *Journal of the American College of Cardiology*, 47(11):2141–2151.

[60] Graatsma EM, Jacod BC, Egmond LAJ Van, Mulder EJH, Visser GHA (2009) Fetal electrocardiography: Feasibility of long-term fetal heart rate recordings. *BJOG: An International Journal of Obstetrics and Gynaecology*, 116(2):334–338. https://doi.org/10.1111/j.1471-0528.2008.01951.x

[61] Mukkamala R, Hahn JO, Inan OT, Mestha LK, Kim CS, Toreyin H, Kyal S (2015) Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. *IEEE Transactions on Biomedical Engineering*, 62(8):1879–1901.

https://doi.org/10.1109/TBME.2015.2441951

[62] Patterson RP (1989) Fundamentals of impedance cardiography. *IEEE Engineering in Medicine and Biology magazine*, 8(1):35–38.

[63] Criée CP, Sorichter S, Smith HJ, Kardos P, Merget R, Heise D, Berdel D, Köhler D, Magnussen H, Marek W (2011) Body plethysmography–its principles and clinical use. *Respiratory medicine*, 105(7):959–971.

[64] Choi Y, Zhang Q, Ko S (2013) Noninvasive cuffless blood pressure estimation using pulse transit time and Hilbert–Huang transform. *Computers & Electrical Engineering*, 39(1):103–111.

[65] Kim C-S, Carek AM, Inan OT, Mukkamala R, Hahn J-O (2018) Ballistocardiogram-based approach to cuffless blood pressure monitoring: proof of concept and potential challenges. *IEEE Transactions on Biomedical Engineering*, 65(11):2384–2391.

[66] Kim BS, Yoo SK (2006) Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE transactions on biomedical engineering*, 53(3):566–568.

[67] Abbas AK, Bassam R (2009) Phonocardiography Signal Processing. *Synthesis Lectures on Biomedical Engineering*, 31(1):1–189. https://doi.org/10.2200/S00187ED1V01Y200904BME031

[68] Jain PK, Tiwari AK (2014) Heart monitoring systems—A review. *Computers in biology and medicine*, 54:1–13.

[69] Inan OT, Migeotte P-F, Park K-S, Etemadi M, Tavakolian K, Casanella R, Zanetti J, Tank J, Funtova I, Prisk GK (2014) Ballistocardiography and seismocardiography: A review of recent advances. *IEEE journal of biomedical and health informatics*, 19(4):1414–1427.

[70] Hersek S, Semiz B, Shandhi MMH, Orlandic L, Inan OT (2019) A Globalized Model for Mapping Wearable Seismocardiogram Signals to Whole-Body Ballistocardiogram Signals Based on Deep Learning. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1296–1309.

[71] Yang C, Aranoff ND, Green P, Tavassolian N (2020) Classification of Aortic Stenosis Using Time-Frequency Features From Chest Cardio-Mechanical Signals. *IEEE Trans. Biomed. Engineering*, 67(6):1672–1683.

[72] Yang C, Ojha BD, Aranoff ND, Green P, Tavassolian N (2020) Classification of aortic stenosis using conventional machine learning and deep learning methods based on multi-dimensional cardio-mechanical signals. *Scientific Reports*, 10(1):1–11. https://doi.org/10.1038/s41598-020-

74519-6

[73] Yang C, Tavassolian N (2018) Combined Seismo-and Gyro-Cardiography: A More Comprehensive Evaluation of Heart-Induced Chest Vibrations (Cover Article). *IEEE Journal of Biomedical and Health Informatics*, 22(5):1466–1475. https://doi.org/10.1109/JBHI.2017.2764798

[74] Yang C, Tang S, Tavassolian N (2017) Utilizing Gyroscopes Towards the Automatic Annotation of Seismocardiograms. *IEEE Sensors Journal*, 17(7):2129–2136. https://doi.org/10.1109/JSEN.2017.2663420

[75] Zanetti JM, Tavakolian K (2013) Seismocardiography: Past, present and future. *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, :7004–7007.

[76] Yang C, Tavassolian N (2018) Pulse transit time measurement using seismocardiogram, photoplethysmogram, and acoustic recordings: Evaluation and comparison. *IEEE Journal of Biomedical and Health Informatics*, 22(3):733–740. https://doi.org/10.1109/JBHI.2017.2696703

[77] Carek AM, Conant J, Joshi A, Kang H, Inan OT (2017) SeismoWatch: wearable cuffless blood pressure monitoring using pulse transit time. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 1(3):1–16.

[78] Rienzo M Di, Vaini E, Lombardi P (2015) Use of seismocardiogram for the beat-to-beat assessment of the Pulse Transit Time: A pilot study. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, :7184–7187.

[79] Taebi A, Solar BE, Bomar AJ, Sandler RH, Mansy HA (2019) Recent advances in seismocardiography. *Vibration*, 2(1):64–86.

[80] Wang EJ, Zhu J, Jain M, Lee T-J, Saba E, Nachman L, Patel SN (2018) Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, :1–9.

[81] Ganti VG, Carek A, Nevius BN, Heller J, Etemadi M, Inan O (2020) Wearable Cuff-less Blood Pressure Estimation at Home via Pulse Transit Time. *IEEE Journal of Biomedical and Health Informatics*,

[82] Tadi MJ, Lehtonen E, Saraste A, Tuominen J, Koskinen J, Teräs M, Airaksinen J, Pänkäälä

M, Koivisto T (2017) Gyrocardiography: A new non-invasive monitoring method for the assessment of cardiac mechanics and the estimation of hemodynamic variables. *Scientific reports*, 7(1):1–11.

[83] Khosrow-Khavar F, Tavakolian K, Blaber A, Menon C (2017) Automatic and robust delineation of the fiducial points of the seismocardiogram signal for noninvasive estimation of cardiac time intervals. *IEEE Transactions on Biomedical Engineering*, 64(8):1701–1710. https://doi.org/10.1109/TBME.2016.2616382

[84] Tadi MJ, Lehtonen E, Hurnanen T, Koskinen J, Eriksson J, Pänkäälä M, Teräs M, Koivisto T (2016) A real-time approach for heart rate monitoring using a Hilbert transform in seismocardiograms. *Physiological measurement*, 37(11):1885.

[85] Choudhary T, Sharma LN, Bhuyan MK (2019) Automatic detection of aortic valve opening using seismocardiography in healthy individuals. *IEEE Journal of Biomedical and Health Informatics*, 23(3):1032–1040. https://doi.org/10.1109/JBHI.2018.2829608

[86] Cheitlin MD, Gertz EW, Brundage BH, Carlson CJ, Quash JA, Bode Jr RS (1979) Rate of progression of severity of valvular aortic stenosis in the adult. *American heart journal*, 98(6):689–700.

[87] Tschumper M, Weber L, Rickli H, Seidl S, Brenner R, Buser M, Ehl NF, Jäger-Rhomberg F, Ammann P, Maeder MT (2021) Corrected QT Interval in Severe Aortic Stenosis: Clinical and Hemodynamic Correlates and Prognostic Impact. *American Journal of Medicine*, 134(2):267–277. https://doi.org/10.1016/j.amjmed.2020.05.035

[88] Jung J, Heisel A, Butz B, Fries R, Schieffer H, Tscholl D, Schäfers H (1997) Factors influencing heart rate variability in patients with severe aortic valve disease. *Clinical cardiology*, 20(4):341–344.

[89] Benitez D, Gaydecki PA, Zaidi A, Fitzpatrick AP (2001) The use of the Hilbert transform in ECG signal analysis. *Computers in Biology and Medicine*, 31(5):399–406. https://doi.org/10.1016/S0010-4825(01)00009-9

[90] Pan J, Tompkins WJ (1985) A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236. https://doi.org/10.1109/TBME.1985.325532

[91] Shaffer F, Ginsberg JP (2017) An overview of heart rate variability metrics and norms.

*Frontiers in public health*, 5:258.

[92] Sieciński S, Kostka PS, Tkacz EJ (2020) Heart Rate Variability Analysis on Electrocardiograms, Seismocardiograms and Gyrocardiograms on Healthy Volunteers. *Sensors*, 20(16):4522.

[93] Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS (2006) Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051.

[94] Brennan M, Palaniswami M, Kamen P (2001) Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE transactions on biomedical engineering*, 48(11):1342–1347.

[95] Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ, Coumel P, Fallen EL, Kennedy HL, Kleiger RE (1996) Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology.

[96] Murphy KP (2012) https://mitpress.mit.edu/books/machine-learning-1Machine Learning A Probabilistic Perspective.

[97] Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, :785–794.

[98] Ogunleye A, Wang QG (2020) XGBoost Model for Chronic Kidney Disease Diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6):2131–2140. https://doi.org/10.1109/TCBB.2019.2911071

[99] Chicco D, Oneto L (2021) An Enhanced Random Forests Approach to Predict Heart Failure from Small Imbalanced Gene Expression Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6):2759–2765. https://doi.org/10.1109/TCBB.2020.3041527

[100] Shokouhmand S, Rahman MM, Faezipour M, Bhatt S (2023) Adventitious Pulmonary Sound Detection Using XGBoost. *2023 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*,

[101] Shokouhmand S, Rahman MM, Faezipour M, Bhatt S (2023) Abnormality Detection in Lung Sounds Using Feature Augmentation. *The 9th International Conference on Health Informatics and Medical Systems*,

[102] Sun B-Y, Zhu Z-H, Li J, Linghu B (2010) Combined feature selection and cancer prognosis using support vector machine regression. *IEEE/ACM transactions on computational biology and bioinformatics*, 8(6):1671–1677.

[103] Iftikhar Z, Lahdenoja O, Tadi MJ, Hurnanen T, Vasankari T, Kiviniemi T, Airaksinen J, Koivisto T, Pänkäälä M (2018) Multiclass classifier based cardiovascular condition detection using smartphone mechanocardiography. *Scientific reports*, 8(1):1–14.

[104] Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, :2951–2959.

[105] Lundberg S, Lee S-I (2017) A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*,

[106] Lahdenoja O, Koivisto T, Tadi MJ, Iftikhar Z, Hurnanen T, Vasankari T, Kiviniemi T, Airaksinen J, Pankaala M (2017) A smartphone-only solution for detecting indications of acute myocardial infarction. *2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017*, :197–200. https://doi.org/10.1109/BHI.2017.7897239

[107] Jafari Tadi M, Mehrang S, Kaisti M, Lahdenoja O, Hurnanen T, Jaakkola J, Jaakkola S, Vasankari T, Kiviniemi T, Airaksinen J, Knuutila T, Lehtonen E, Koivisto T, Pänkäälä M (2019) Comprehensive Analysis of Cardiogenic Vibrations for Automated Detection of Atrial Fibrillation Using Smartphone Mechanocardiograms. *IEEE Sensors Journal*, 19(6):2230–2242. https://doi.org/10.1109/JSEN.2018.2882874

[108] Lahdenoja O, Hurnanen T, Iftikhar Z, Nieminen S, Knuutila T, Saraste A, Kiviniemi T, Vasankari T, Airaksinen J, Pänkäälä M (2017) Atrial fibrillation detection via accelerometer and gyroscope of a smartphone. *IEEE Journal of Biomedical and Health Informatics*, 22(1):108–118.

[109] Mallow PJ, Chen J, Moore M, Gunnarsson C, Rizzo JA (2019) Incremental direct healthcare expenditures of valvular heart disease in the USA. *Journal of Comparative Effectiveness Research*, 8(11):879–887.

[110] Thaden JJ, Tsang MY, Ayoub C, Padang R, Nkomo VT, Tucker SF, Cassidy CS, Bremer M, Kane GC, Pellikka PA (2017) Association Between Echocardiography Laboratory Accreditation and the Quality of Imaging and Reporting for Valvular Heart Disease. *Circulation. Cardiovascular imaging*, 10(8):e006140. https://doi.org/10.1161/CIRCIMAGING.117.006140

[111] Birrane JP, Lim ZL, Liew CH, Rosseel L, Heerey A, Coleman K, Gallagher J, Mylotte D, McEvoy JW (2022) A survey of general practitioners' knowledge and clinical practice in relation to valvular heart disease. *Irish Journal of Medical Science (1971-)*, 191(2):777–784.

[112] Cohen-Shelly M, Attia ZI, Friedman PA, Ito S, Essayagh BA, Ko W-Y, Murphree DH, Michelena HI, Enriquez-Sarano M, Carter RE (2021) Electrocardiogram screening for aortic valve stenosis using artificial intelligence. *European heart journal*, 42(30):2885–2896.

[113] Kwon J, Lee SY, Jeon K, Lee Y, Kim K, Park J, Oh B, Lee M (2020) Deep learning–based algorithm for detecting aortic stenosis using electrocardiography. *Journal of the American Heart Association*, 9(7):e014717.

[114] Elias P, Poterucha TJ, Rajaram V, Moller LM, Rodriguez V, Bhave S, Hahn RT, Tison G, Abreau SA, Barrios J, Torres JN, Hughes JW, Perez M V., Finer J, Kodali S, Khalique O, Hamid N, Schwartz A, Homma S, Kumaraiah D, Cohen DJ, Maurer MS, Einstein AJ, Nazif T, Leon MB, Perotte AJ (2022) Deep Learning Electrocardiographic Analysis for Detection of Left-Sided Valvular Heart Disease. *Journal of the American College of Cardiology*, 80(6):613–626. https://doi.org/10.1016/j.jacc.2022.05.029

[115] Gharehbaghi A, Linden M, Babic A (2019) An artificial intelligent-based model for detecting systolic pathological patterns of phonocardiogram based on time-growing neural network. *Applied Soft Computing*, 83:105615.

[116] Ghosh SK, Tripathy RK, Ponnalagu RN, Pachori RB (2019) Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features. *IEEE Sensors Letters*, 3(12):1–4.

[117] Alqudah AM (2019) Towards classifying non-segmented heart sound records using instantaneous frequency based features. *Journal of medical engineering & technology*, 43(7):418–430.

[118] Nizam NB, Nuhash SISK, Hasan T (2022) Hilbert-envelope features for cardiac disease classification from noisy phonocardiograms. *Biomedical Signal Processing and Control*, 78:103864. https://doi.org/10.1016/j.bspc.2022.103864

[119] Ahmad MS, Mir J, Ullah MO, Shahid MLUR, Syed MA (2019) An efficient heart murmur recognition and cardiovascular disorders classification system. *Australasian Physical and Engineering Sciences in Medicine*, 42(3):733–743. https://doi.org/10.1007/s13246-019-00778-x

[120] Yaseen, Son GY, Kwon S (2018) Classification of heart sound signal using multiple features. *Applied Sciences (Switzerland)*, 8(12):2344. https://doi.org/10.3390/app8122344

[121] Alkhodari M, Fraiwan L (2021) Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. *Computer Methods and Programs in Biomedicine*, 200:105940.

[122] Khan JS, Kaushik M, Chaurasia A, Dutta MK, Burget R (2022) Cardi-Net: A deep neural network for classification of cardiac disease using phonocardiogram signal. *Computer Methods and Programs in Biomedicine*, 219:106727.

[123] McLane I, Emmanouilidou D, West JE, Elhilali M (2021) Design and Comparative Performance of a Robust Lung Auscultation System for Noisy Clinical Settings. *IEEE Journal of Biomedical and Health Informatics*, 25(7):2583–2594. https://doi.org/10.1109/JBHI.2021.3056916

[124] Kalinauskienė E, Razvadauskas H, Morse DJ, Maxey GE, Naudžiūnas A (2019) A comparison of electronic and traditional stethoscopes in the heart auscultation of obese patients. *Medicina*, 55(4):94.

[125] Shokouhmand A, Aranoff ND, Driggin E, Green P, Tavassolian N (2021) Efficient Detection of Aortic Stenosis Using Morphological Characteristics of Cardiomechanical Signals and Heart Rate Variability Parameters. *Scientific Reports*, 11(1)

[126] Shokouhmand A, Yang C, Aranoff ND, Driggin E, Green P, Tavassolian N (2021) Mean Pressure Gradient Prediction Based on Chest Angular Movements and Heart Rate Variability Parameters. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*,

[127] Coviello JS (2013) Auscultation skills: Breath & heart sounds.

[128] Gupta P, Wen H, Francesco L Di, Ayazi F (2021) Detection of pathological mechano-acoustic signatures using precision accelerometer contact microphones in patients with pulmonary disorders. *Scientific Reports*, 11(1):1–12.

[129] Gupta P, Moghimi MJ, Jeong Y, Gupta D, Inan OT, Ayazi F (2020) Precision wearable accelerometer contact microphones for longitudinal monitoring of mechano-acoustic cardiopulmonary signals. *npj Digital Medicine*, 3(1):1–8. https://doi.org/10.1038/s41746-020-

0225-7

[130] Rennert NJ, Morris R, Barrere CC (2004) How to cope with scopes: Stethoscope Selection and use with hearing aids and Cls. *Hearing Review*, 11(2):34–41.

[131] Flandrin P, Goncalves P, Rilling G (2004) Detrending and denoising with empirical mode decompositions. *2004 12th European signal processing conference*, :1581–1584.

[132] Shokouhmand A, Wen H, Khan S, Puma JA, Patel A, Green P, Ayazi F, Tavassolian N (2023) Diagnosis of Peripheral Artery Disease Using Backflow Abnormalities in Proximal Recordings of Accelerometer Contact Microphone (ACM) (Cover Article). *IEEE Journal of Biomedical and Health Informatics*, 27(1):274–285. https://doi.org/10.1109/JBHI.2022.3218595

[133] Savitzky A, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639. https://doi.org/10.1021/ac60214a047

[134] Li J, Deng L, Haeb-Umbach R, Gong Y (2016) Fundamentals of speech recognition. *Robust Automatic Speech Recognition*, https://doi.org/10.1016/b978-0-12-802398-3.00002-7

[135] Rawat W, Wang Z (2017) Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449.

[136] Gheshlaghi SH, Nok Enoch Kan C, Ye DH (2021) Breast Cancer Histopathological Image Classification with Adversarial Image Synthesis. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, :3387–3390. https://doi.org/10.1109/EMBC46164.2021.9630678

[137] Tian C, Fei L, Zheng W, Xu Y, Zuo W, Lin C-W (2020) Deep learning on image denoising: An overview. *Neural Networks*, 131:251–275.

[138] Rahman MM, Shokouhmand S, Faezipour M, Bhatt S (2022) Attentional Convolutional Neural Network for Automating Pathological Lung Auscultations Using Respiratory Sounds. *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*,

[139] Guo Y, Liu Y, Georgiou T, Lew MS (2018) A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval*, 7(2):87–93.

[140] Heidari Gheshlaghi S, Kan CNE, Gilat-Schmidt T, Ye DH (2022) Age encoded adversarial learning for pediatric CT segmentation.

[141] To T, Gheshlaghi SH, Ye DH (2022) Deep learning for breast cancer classification of deep ultraviolet fluorescence images toward intra-operative margin assessment. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, :1891–1894.

[142] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S (2020) An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*,

[143] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-Decem:5999–6009.

[144] Guo J, Han K, Wu H, Tang Y, Chen X, Wang Y, Xu C (2022) Cmt: Convolutional neural networks meet vision transformers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, :12175–12185.

[145] Hendrycks D, Gimpel K (2016) Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*,

[146] Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning*, :448–456.

[147] Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*,

[148] Powell-Wiley TM, Poirier P, Burke LE, Després J-P, Gordon-Larsen P, Lavie CJ, Lear SA, Ndumele CE, Neeland IJ, Sanders P (2021) Obesity and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation*, 143(21):e984–e1010.

[149] Kaltoft M, Langsted A, Nordestgaard BG (2020) Obesity as a Causal Risk Factor for Aortic Valve Stenosis. *Journal of the American College of Cardiology*, 75(2):163–176. https://doi.org/10.1016/j.jacc.2019.10.050

[150] Fontecave-Jallon J, Fojtik K, Rivet B (2019) Is there an Optimal Localization of Cardio-microphone Sensors for Phonocardiogram Analysis? *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, :3249–3252. https://doi.org/10.1109/EMBC.2019.8857681

[151] Javaid AQ, Ashouri H, Dorier A, Etemadi M, Heller JA, Roy S, Inan OT (2017) Quantifying and reducing motion artifacts in wearable seismocardiogram measurements during walking to assess left ventricular health. *IEEE Transactions on Biomedical Engineering*, 64(6):1277–1286. https://doi.org/10.1109/TBME.2016.2600945

[152] Lee ES, Ha EL, Yang EH, Vorobiof G (2014) Diagnosis and Management of Valvular Heart Disease. *Hospital Medicine Clinics*, 3(3):483–500. https://doi.org/10.1016/j.ehmc.2014.03.001

[153] He W, Xiao H, Liu X (2012) Numerical simulation of human systemic arterial hemodynamics based on a transmission line model and recursive algorithm. *Journal of Mechanics in Medicine and Biology*, 12(1):1250020. https://doi.org/10.1142/S0219519411004587

[154] Drake R, Vogl AW, Mitchell AWM (2009) Gray's anatomy for students E-book.

[155] Salman AH, Ahmadi N, Mengko R, Langi AZR, Mengko TLR (2015) Performance comparison of denoising methods for heart sound signal. *2015 international symposium on intelligent signal processing and communication systems (ISPACS)*, :435–440.

[156] Hussain M, Bird JJ, Faria DR (2018) A study on cnn transfer learning for image classification. *UK Workshop on computational Intelligence*, :191–202.

[157] Sultana F, Sufian A, Dutta P (2020) Evolution of image segmentation using deep convolutional neural network: a survey. *Knowledge-Based Systems*, 201:106062.

[158] Xie J, Xu L, Chen E (2012) Image denoising and inpainting with deep neural networks. *Advances in neural information processing systems*, 25

[159] Yu F, Xiu X, Li Y (2022) A Survey on Deep Transfer Learning and Beyond. *Mathematics*, 10(19):270–279. https://doi.org/10.3390/math10193619

[160] He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, :770–778.

[161] Santurkar S, Tsipras D, Ilyas A, Madry A (2018) How does batch normalization help optimization? *Advances in Neural Information Processing Systems*, 2018-Decem:2483–2493.

[162] Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, DePalma SM, Gidding S, Jamerson KA, Jones DW (2018) 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the

American College of Cardiology/American Heart Association Task Force on Clinical Pr. *Journal of the American College of Cardiology*, 71(19):e127–e248.

[163] Centers for Disease Control and Prevention (2021) Estimated Hypertension Prevalence, Treatment, and Control Among U.S. Adults. https://www.cdc.gov/chronicdisease/programs-impact/pop/high-blood-pressure.htm

[164] Stamler J, Neaton JD, Wentworth DN (1989) Blood pressure (systolic and diastolic) and risk of fatal coronary heart disease. *Hypertension*, 13(5_supplement):I2.

[165] Lawes CMM, Bennett DA, Feigin VL, Rodgers A (2004) Blood pressure and stroke: an overview of published reviews. *Stroke*, 35(3):776–785.

[166] Levin NW, Kotanko P, Eckardt K-U, Kasiske BL, Chazot C, Cheung AK, Redon J, Wheeler DC, Zoccali C, London GM (2010) Blood pressure in chronic kidney disease stage 5D—report from a Kidney Disease: Improving Global Outcomes controversies conference. *Kidney international*, 77(4):273–284.

[167] Overwyk KJ, Dehmer SP, Roy K, Maciosek M V, Hong Y, Baker-Goering MM, Loustalot F, Singleton C-M, Ritchey MD (2019) Modeling the health and budgetary impacts of a team-based hypertension care intervention that includes pharmacists. *Medical care*, 57(11):882.

[168] Viera AJ, Cohen LW, Mitchell CM, Sloane PD (2008) Use of home blood pressure monitoring by hypertensive patients in primary care: survey of a practice-based research network cohort. *The Journal of Clinical Hypertension*, 10(4):280–286.

[169] Nessler K, Krztoń-Królewiecka A, Suska A, Mann MR, Nessler MB, Windak A (2021) The quality of patients' self-blood pressure measurements: a cross-sectional study. *BMC Cardiovascular Disorders*, 21:1–8.

[170] El-Hajj C, Kyriacou PA (2020) A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. *Biomedical Signal Processing and Control*, 58:101870.

[171] González S, Hsieh W-T, Chen TP-C (2023) A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram. *Scientific Data*, 10(1):149.

[172] Martínez G, Howard N, Abbott D, Lim K, Ward R, Elgendi M (2018) Can photoplethysmography replace arterial blood pressure in the assessment of blood pressure? *Journal*

*of clinical medicine*, 7(10):316.

[173] Liu D, Görges M, Jenkins SA (2012) University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research. *Anesthesia & Analgesia*, 114(3):584–589.

[174] Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, Moody B, Szolovits P, Anthony Celi L, Mark RG (2016) MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

[175] Johnson A, Pollard T, Mark R (2019) Mimic-iii clinical database demo (version 1.4). *PhysioNet*, 10:C2HM2Q.

[176] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23)https://doi.org/10.1161/01.cir.101.23.e215

[177] Liang Y, Chen Z, Liu G, Elgendi M (2018) A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. *Scientific data*, 5(1):1–7.

[178] Maqsood S, Xu S, Springer M, Mohawesh R (2021) A benchmark study of machine learning for analysis of signal feature extraction techniques for blood pressure estimation using photoplethysmography (PPG). *IEEE Access*, 9:138817–138833.

[179] Duan K, Qian Z, Atef M, Wang G (2016) A feature exploration methodology for learning based cuffless blood pressure measurement using photoplethysmography. *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, :6385–6388.

[180] Mousavi SS, Firouzmand M, Charmi M, Hemmati M, Moghadam M, Ghorbani Y (2019) Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method. *Biomedical Signal Processing and Control*, 47:196–206.

[181] Wang W, Mohseni P, Kilgore KL, Najafizadeh L (2021) Cuff-less blood pressure estimation from photoplethysmography via visibility graph and transfer learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2075–2085.

[182] Slapničar G, Mlakar N, Luštrek M (2019) Blood pressure estimation from

photoplethysmogram using a spectro-temporal deep neural network. *Sensors*, 19(15):3420.

[183] Huang B, Chen W, Lin C-L, Juang C-F, Wang J (2022) MLP-BP: A novel framework for cuffless blood pressure measurement with PPG and ECG signals based on MLP-Mixer neural networks. *Biomedical Signal Processing and Control*, 73:103404.

[184] Aguirre N, Grall-Maës E, Cymberknop LJ, Armentano RL (2021) Blood pressure morphology assessment from photoplethysmogram and demographic information using deep learning with attention mechanism. *Sensors*, 21(6):2167.

[185] Fallow BA, Tarumi T, Tanaka H (2013) Influence of skin type and wavelength on light wave reflectance. *Journal of clinical monitoring and computing*, 27:313–317.

[186] Zhang Y, Liu B, Zhang Z (2015) Combining ensemble empirical mode decomposition with spectrum subtraction technique for heart rate monitoring using wrist-type photoplethysmography. *Biomedical Signal Processing and Control*, 21:119–125.

[187] Wang Y-J, Chen C-H, Sue C-Y, Lu W-H, Chiou Y-H (2018) Estimation of blood pressure in the radial artery using strain-based pulse wave and photoplethysmography sensors. *Micromachines*, 9(11):556.

[188] Wang Y-J, Chen T-Y, Tsai M-C, Wu CH (2016) Noninvasive blood pressure monitor using strain gauges, a fastening band, and a wrist elasticity model. *Sensors and Actuators A: Physical*, 252:198–208.

[189] Pang Y, Yang Z, Han X, Jian J, Li Y, Wang X, Qiao Y, Yang Y, Ren T-L (2018) Multifunctional mechanical sensors for versatile physiological signal detection. *ACS applied materials & interfaces*, 10(50):44173–44182.

[190] Suster M, Guo J, Chaimanonart N, Ko WH, Young DJ (2006) A high-performance MEMS capacitive strain sensing system. *Journal of Microelectromechanical Systems*, 15(5):1069–1077.

[191] Bagha S, Shaw L (2011) A real time analysis of PPG signal for measurement of SpO2 and pulse rate. *International journal of computer applications*, 36(11):45–50.

[192] Dawber TR, THomas Jr HE, McNamara PM (1973) Characteristics of the dicrotic notch of the arterial pulse wave in coronary heart disease. *Angiology*, 24(4):244–255.

[193] Gupta S, Singh A, Sharma A (2022) Dynamic large artery stiffness index for cuffless blood pressure estimation. *IEEE Sensors Letters*, 6(3):1–4.

[194] Stergiou GS, Alpert B, Mieke S, Asmar R, Atkins N, Eckert S, Frick G, Friedman B, Graßl T, Ichikawa T (2018) A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement. *Hypertension*, 71(3):368–374.

[195] Ternacle J, Clavel M-A (2020) Assessment of aortic stenosis severity: a multimodality approach. *Cardiology clinics*, 38(1):13–22.

[196] Inan OT, Baran Pouyan M, Javaid AQ, Dowling S, Etemadi M, Dorier A, Heller JA, Bicen AO, Roy S, Marco T De (2018) Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients. *Circulation: Heart Failure*, 11(1):e004313.

[197] Golovchiner G, Glikson M, Swissa M, Sela Y, Abelow A, Morelli O, Beker A (2022) Automated detection of atrial fibrillation based on vocal features analysis. *Journal of Cardiovascular Electrophysiology*, 33(8):1647–1654. https://doi.org/10.1111/jce.15595

[198] Yang C, Aranoff ND, Green P, Tavassolian N (2018) A Binary Classification of Cardiovascular Abnormality Using Time-Frequency Features of Cardio-mechanical Signals. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2018:5438–5441. https://doi.org/10.1109/EMBC.2018.8513644

[199] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A (2018) Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 2018-Decem:6638–6648.

[200] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30:3146–3154.

[201] Ashrafimoghari V, Suchow JW (2022) A game-theoretic model of the consumer behavior under pay-what-you-want pricing strategy. *arXiv preprint arXiv:2207.08923*,

[202] Yancy CW, Vorovich E, Chuzi S (2021) Heart Failure With Reduced Ejection Fraction. *Sex Differences in Cardiac Diseases: Pathophysiology, Presentation, Diagnosis and Management*, 324(5):185–200. https://doi.org/10.1016/B978-0-12-819369-3.00022-8

[203] Fonarow GC, Hsu JJ (2016) Left Ventricular Ejection Fraction. *JACC: Heart Failure*, 4(6):511–513. https://doi.org/10.1016/j.jchf.2016.03.021

[204] Hayley BD, Burwash IG (2012) Heart failure with normal left ventricular ejection fraction: Role of echocardiography. *Current Opinion in Cardiology*, 27(2):169–180. https://doi.org/10.1097/HCO.0b013e32834fe8df

[205] Frauenrath T, Hezel F, Heinrichs U, Kozerke S, Utting JF, Kob M, Butenweg C, Boesiger P, Niendorf T (2009) Feasibility of cardiac gating free of interference with electro-magnetic fields at 1.5 Tesla, 3.0 Tesla and 7.0 Tesla using an MR-stethoscope. *Investigative radiology*, 44(9):539–547.

[206] Varghees VN, Ramachandran KI (2017) Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope. *IEEE Sensors Journal*, 17(12):3861–3872.

[207] Klum M, Urban M, Tigges T, Pielmus A-G, Feldheiser A, Schmitt T, Orglmeister R (2020) Wearable cardiorespiratory monitoring employing a multimodal digital patch stethoscope: estimation of ECG, PEP, LVET and respiration using a 55 mm single-lead ECG and phonocardiogram. *Sensors*, 20(7):2033.

[208] Gupta P, Moghimi MJ, Jeong Y, Gupta D, Inan OT, Ayazi F (2020) Precision wearable accelerometer contact microphones for longitudinal monitoring of mechano-acoustic cardiopulmonary signals. *NPJ digital medicine*, 3(1):1–8.

[209] Garrard Jr CL, Weissler AM, Dodge HT (1970) The relationship of alterations in systolic time intervals to ejection fraction in patients with cardiac disease. *Circulation*, 42(3):455–462.

[210] Oppenheim A van, Schafer R, Stockham T (1968) Nonlinear filtering of multiplied and convolved signals. *IEEE transactions on audio and electroacoustics*, 16(3):437–466.

[211] Park MK (2014) Pediatric cardiology for practitioners E-Book.

[212] Tamura T (2019) Current progress of photoplethysmography and SPO2 for health monitoring. *Biomedical Engineering Letters*, 9(1):21–36. https://doi.org/10.1007/s13534-019-00097-w

[213] Lázaro J, Gil E, Bailón R, Laguna P (2011) Deriving respiration from the pulse photoplethysmographic signal. *2011 Computing in Cardiology*, :713–716.

[214] Lin W-H, Wu D, Li C, Zhang H, Zhang Y-T (2014) Comparison of heart rate variability from PPG with that from ECG. *The International Conference on Health Informatics: ICHI 2013, Vilamoura, Portugal on 7-9 November, 2013*, :213–215.

[215] Xing X, Sun M (2016) Optical blood pressure estimation with photoplethysmography and FFT-based neural networks. *Biomedical optics express*, 7(8):3007–3020.

[216] Bent B, Goldstein BA, Kibbe WA, Dunn JP (2020) Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ digital medicine*, 3(1):18.

[217] Nyboer J, Kreider MM, Hannapel L (1950) Electrical impedance plethysmography: a physical and physiologic approach to peripheral vascular study. *Circulation*, 2(6):811–821.

[218] Whitney RJ (1953) The measurement of volume changes in human limbs. *The Journal of Physiology*, 121(1):1–27. https://doi.org/10.1113/jphysiol.1953.sp004926

[219] Campinho P, Vilfan A, Vermot J (2020) Blood flow forces in shaping the vascular system: a focus on endothelial cell behavior. *Frontiers in Physiology*, 11:552.

[220] Shokouhmand A, Wen H, Khan S, Puma JA, Patel A, Green P, Ayazi F, Tavassolian N Diagnosis of Coexisting Valvular Heart Diseases Using Image-to-Sequence Translation of Contact Microphone Recordings. *IEEE Transactions on Biomedical Engineering*,

[221] Sang B, Wen H, Gupta P, Shokouhmand A, Khan S, Puma JA, Patel A, Green P, Tavassolian N, Ayazi F (2022) Detection of Normal and Paradoxical Splitting in Second Heart Sound (S2) using a Wearable Accelerometer Contact Microphone. *2022 IEEE Sensors*, :1–4.

[222] Shokouhmand A, Wen H, Khan S, Puma JA, Patel A, Green P, Ayazi F, Tavassolian N (2022) Detection of Left Ventricular Ejection Fraction Abnormality Using Fusion of Acoustic and Biopotential Characteristics of Precordium. *2022 IEEE Sensors*, :1–4.

[223] Etemadi M, Inan OT (2018) Wearable ballistocardiogram and seismocardiogram systems for health and performance. *Journal of Applied Physiology*, 124(2):452–461.

[224] Ramírez J, Rodriquez D, Urbina AD, Cardenas AM, Lipomi DJ (2019) Combining high sensitivity and dynamic range: wearable thin-film composite strain sensors of graphene, ultrathin palladium, and PEDOT: PSS. *ACS applied nano materials*, 2(4):2222–2229.

[225] Sensing S Consensys ECG Development Kits. https://shimmersensing.com/product/consensys-ecg-development-kits/

[226] Kaplan BM, Langendorf R, Lev M, Pick A (1973) Tachycardia-bradycardia syndrome (so-called “sick sinus syndrome”): pathology, mechanisms and treatment. *The American journal of cardiology*, 31(4):497–508.

[227] Berntson GG, Cacioppo JT, Quigley KS (1993) Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology*, 30(2):183–196.

[228] Brinkmann AE, Press SA, Helmert E, Hautzinger M, Khazan I, Vagedes J (2020) Comparing effectiveness of HRV-biofeedback and mindfulness for workplace stress reduction: A randomized controlled trial. *Applied psychophysiology and biofeedback*, 45:307–322.

[229] Reed MJ, Robertson CE, Addison PS (2005) Heart rate variability measurements and the prediction of ventricular arrhythmias. *Qjm*, 98(2):87–95.

[230] Huikuri H V (1995) Heart rate variability in coronary artery disease. *Journal of internal medicine*, 237(4):349–357.

[231] Shokouhmand A, Ayazi F, Ebadi N (2023) Fingertip Strain Plethysmography: Representation of Pulse Information based on Vascular Vibration. *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, :1–5.

[232] Choudhary T, Das M, Sharma LN, Bhuyan MK (2021) Analyzing seismocardiographic approach for heart rate variability measurement. *Biomedical Signal Processing and Control*, 68:102793.

[233] Hoog Antink C, Mai Y, Peltokangas M, Leonhardt S, Oksala N, Vehkaoja A (2021) Accuracy of heart rate variability estimated with reflective wrist-PPG in elderly vascular patients. *Scientific reports*, 11(1):8123.

[234] (WHO) WHO (2020) WHO reveals leading causes of death and disability worldwide: 2000-2019. *(WHO), World Health Organization*, https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019

[235] Franklin SS, Khan SA, Wong ND, Larson MG, Levy D (1999) Is pulse pressure useful in predicting risk for coronary heart disease? The Framingham Heart Study. *Circulation*, 100(4):354–360.

[236] Arakawa T (2018) Recent research and developing trends of wearable sensors for detecting

blood pressure. *Sensors*, 18(9):2772.

[237] Kim C-S, Carek AM, Mukkamala R, Inan OT, Hahn J-O (2015) Ballistocardiogram as proximal timing reference for pulse transit time measurement: Potential for cuffless blood pressure monitoring. *IEEE Transactions on Biomedical Engineering*, 62(11):2657–2664.

[238] Chen W, Kobayashi T, Ichikawa S, Takeuchi Y, Togawa T (2000) Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration. *Medical and Biological Engineering and Computing*, 38:569–574.

[239] Tang Z, Tamura T, Sekine M, Huang M, Chen W, Yoshida M, Sakatani K, Kobayashi H, Kanaya S (2016) A chair–based unobtrusive cuffless blood pressure monitoring system based on pulse arrival time. *IEEE Journal of biomedical and health informatics*, 21(5):1194–1205.

[240] Lee J, Sohn JJ, Park J, Yang SM, Lee S, Kim HC (2018) Novel blood pressure and pulse pressure estimation based on pulse transit time and stroke volume approximation. *BioMedical Engineering Online*, 17(1):1–20. https://doi.org/10.1186/s12938-018-0510-8

[241] Shokouhmand A, Wen H, Khan S, Puma JA, Patel A, Green P, Ayazi F, Ebadi N (2023) Prospects of Cuffless Pulse Pressure Estimation from a Chest-Worn Accelerometer Contact Microphone. *2023 Computing in Cardiology (CinC)*, :1–4.

[242] Kennedy RG (1998) Electronic fetal heart rate monitoring: Retrospective reflections on a twentieth-century technology. *Journal of the Royal Society of Medicine*, 91(5):244–250. https://doi.org/10.1177/014107689809100503

[243] Barnett SB, Maulik D (2001) Guidelines and recommendations for safe use of Doppler ultrasound in perinatal applications. *Journal of Maternal-Fetal Medicine*, 10(2):75–84. https://doi.org/10.1080/jmf.10.2.75.84

[244] Lathauwer L De, Moor B De, Vandewalle J (2000) Fetal electrocardiogram extraction by blind source subspace separation. *IEEE transactions on biomedical engineering*, 47(5):567–572.

[245] Kanjilal PP, Palit S, Saha G (1997) Fetal ECG extraction from single-channel maternal ECG using singular value decomposition. *IEEE Transactions on Biomedical Engineering*, 44(1):51–59. https://doi.org/10.1109/10.553712

[246] Ferrara ER, Widrow B (1982) Fetal Electrocardiogram Enhancement by Time-Sequenced Adaptive Filtering. *IEEE Transactions on Biomedical Engineering*, BME-29(6):458–460.

https://doi.org/10.1109/TBME.1982.324973

[247] Fotiadou E, Vullings R (2020) Multi-Channel Fetal ECG Denoising With Deep Convolutional Neural Networks. *Frontiers in Pediatrics*, 8https://doi.org/10.3389/fped.2020.00508

[248] Fuchs T, Pomorski M, Grobelak K, Tomiałowicz M, Zimmer M (2013) Comment on the article "accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes." *Acta Obstetricia et Gynecologica Scandinavica*, 92(8):992. https://doi.org/10.1111/aogs.12154

[249] Mhajna M, Schwartz N, Levit-Rosen L, Warsof S, Lipschuetz M, Jakobs M, Rychik J, Sohn C, Yagel S (2020) Wireless, remote solution for home fetal and maternal heart rate monitoring. *American Journal of Obstetrics and Gynecology MFM*, 2(2):100101. https://doi.org/10.1016/j.ajogmf.2020.100101

[250] (WHO) WHO (2018) Stillbirth. https://www.who.int/health-topics/stillbirth

[251] Freeman RK, Garite TJ, Nageotte MP, Miller LA (2012) Fetal heart rate monitoring.

[252] Shokouhmand A, Antoine C, Young BK, Tavassolian N (2021) Multi-modal Framework for Fetal Heart Rate Estimation: Fusion of Low-SNR ECG and Inertial Sensors. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, :7166–7169. https://doi.org/10.1109/EMBC46164.2021.9629975

[253] Sameni (2010) A Review of Fetal ECG Signal Processing Issues and Promising Directions. *The Open Pacing, Electrophysiology & Therapy Journal*, 3:4. https://doi.org/10.2174/1876536x01003010004

[254] Kawakita T, Reddy UM, Landy HJ, Iqbal SN, Huang C, Grantz KL (2016) Neonatal complications associated with use of fetal scalp electrode: a retrospective study. *BJOG: An International Journal of Obstetrics & Gynaecology*, 123(11):1797–1803.

[255] Andreotti F, Behar J, Zaunseder S, Oster J, Clifford GD (2016) An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms. *Physiological Measurement*, 37(5):627–648. https://doi.org/10.1088/0967-3334/37/5/627

[256] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components

of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220. https://doi.org/10.1161/01.cir.101.23.e215

[257] Kim T, Cha M, Kim H, Lee JK, Kim J (2017) Learning to discover cross-domain relations with generative adversarial networks. *International Conference on Machine Learning*, :1857–1865.

[258] Lipponen JA, Tarvainen MP (2014) Principal component model for maternal ECG extraction in fetal QRS detection. *Physiological Measurement*, 35(8):1637.

[259] Sameni R, Jutten C, Shamsollahi MB (2008) Multichannel electrocardiogram decomposition using periodic component analysis. *IEEE transactions on biomedical engineering*, 55(8):1935–1940.

[260] Varanini M, Tartarisco G, Billeci L, Macerata A, Pioggia G, Balocchi R (2014) An efficient unsupervised fetal QRS complex detection from abdominal maternal ECG. *Physiological measurement*, 35(8):1607.

[261] Shadaydeh M, Xiao Y, Ward RK (2008) Extraction of fetal ECG using adaptive Volterra filters. *2008 16th European Signal Processing Conference*, :1–5.

[262] Ungureanu GM, Bergmans JWM, Oei SG, Ungureanu A, Wolf W (2009) The event synchronous canceller algorithm removes maternal ECG from abdominal signals without affecting the fetal ECG. *Computers in biology and medicine*, 39(6):562–567.

[263] Abel JDK, Samiappan D, Kumar R, Kumar SP (2019) Multiple sub-filter adaptive noise canceller for fetal ECG extraction. *Procedia Computer Science*, 165:182–188.

[264] Behar J, Johnson A, Clifford GD, Oster J (2014) A comparison of single channel fetal ECG extraction methods. *Annals of biomedical engineering*, 42:1340–1353.

[265] Clifford GD, Silva I, Behar J, Moody GB (2014) Non-invasive fetal ECG analysis. *Physiological measurement*, 35(8):1521.

[266] Sameni R (2008) Extraction of fetal cardiac signals from an array of maternal abdominal recordings.

[267] Sameni R, Shamsollahi MB, Jutten C, Clifford GD (2007) A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Transactions on Biomedical Engineering*, 54(12):2172–2185.

[268] Martens SMM, Rabotti C, Mischi M, Sluijter RJ (2007) A robust fetal ECG detection method

for abdominal recordings. *Physiological measurement*, 28(4):373.

[269] Cerutti S, Baselli G, Civardi S, Ferrazzi E, Marconi AM, Pagani M, Pardi G (1986) Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings.

[270] Ungureanu M, Bergmans JWM, Oei SG, Strungaru R (2007) Fetal ECG extraction during labor using an adaptive maternal beat subtraction technique.

[271] Maria C Di, Liu C, Zheng D, Murray A, Langley P (2014) Extracting fetal heart beats from maternal abdominal recordings: selection of the optimal principal components. *Physiological measurement*, 35(8):1649.

[272] Zhang N, Zhang J, Li H, Mumini OO, Samuel OW, Ivanov K, Wang L (2017) A novel technique for fetal ECG extraction using single-channel abdominal recording. *Sensors*, 17(3):457.

[273] Suganthy M, Joy SI, Sarala B (2019) LabVIEW based non-invasive single channel Fetal Electrocardiogram Extraction. *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, :1–5.

[274] Li G, Zeng X, Zhou X, Liu Q (2012) Robust adaptive fetal heart rate estimation for single-channel abdominal ECG recording. *2012 5th International Conference on BioMedical Engineering and Informatics*, :587–591.

[275] Panigrahy D, Sahu PK (2015) Extraction of fetal electrocardiogram (ECG) by extended state Kalman filtering and adaptive neuro-fuzzy inference system (ANFIS) based on single channel abdominal recording. *Sadhana - Academy Proceedings in Engineering Sciences*, 40(4):1091–1104. https://doi.org/10.1007/s12046-015-0381-7

[276] Niknazar M, Rivet B, Jutten C (2012) Fetal ECG extraction by extended state Kalman filtering based on single-channel recordings. *IEEE Transactions on Biomedical Engineering*, 60(5):1345–1352.

[277] Panigrahy D, Rakshit M, Sahu PK (2015) An efficient method for fetal ECG extraction from single channel abdominal ECG. *2015 international conference on industrial instrumentation and control (ICIC)*, :1083–1088.

[278] Panigrahy D, Sahu PK (2017) Extraction of fetal ECG signal by an improved method using extended Kalman smoother framework from single channel abdominal ECG signal. *Australasian*

*physical & engineering sciences in medicine*, 40(1):191.

[279] Gao P, Chang E-C, Wyse L (2003) Blind separation of fetal ECG from single mixture using SVD and ICA. *Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint*, 3:1418–1422.

[280] He P, Chen X (2016) A method for extracting fetal ECG based on EMD-NMF single channel blind source separation algorithm. *Technology and Health Care*, 24(s1):S17–S26.

[281] He PJ, Chen XM, Liang Y, Zeng HZ (2016) Extraction for fetal ECG using single channel blind source separation algorithm based on multi-algorithm fusion. *MATEC Web of Conferences*, 44:1026.

[282] Mujumdar R, Nadar P, Bondre A, Kulkarni A, Pathak S (2019) Principal Component Analysis (PCA) based single-channel, non-invasive fetal ECG extraction.

[283] Zhang M, Wei G (2020) A complete adaptive method for fetal ECG extraction based on single channel. *Journal of Physics: Conference Series*, 1621(1):12019.

[284] Gurve D, Krishnan S (2019) Separation of fetal-ECG from single-channel abdominal ECG using activation scaled non-negative matrix factorization. *IEEE journal of biomedical and health informatics*, 24(3):669–680.

[285] Krupa AJD, Dhanalakshmi S, Kumar R (2022) Joint time-frequency analysis and non-linear estimation for fetal ECG extraction. *Biomedical Signal Processing and Control*, 75:103569.

[286] Castillo E, Morales DP, García A, Parrilla L, Ruiz VU, Álvarez-Bermejo JA (2018) A clustering-based method for single-channel fetal heart rate monitoring. *PLoS One*, 13(6):e0199308.

[287] Zhong W, Liao L, Guo X, Wang G (2019) Fetal electrocardiography extraction with residual convolutional encoder–decoder networks. *Australasian physical & engineering sciences in medicine*, 42:1081–1089.

[288] Mohebian MR, shahim Vedaei S, Wahid KA, Dinh A, Marateb HR, Tavakolian K (2021) Fetal ECG Extraction from Maternal ECG using Attention-based CycleGAN. *IEEE Journal of Biomedical and Health Informatics*,

[289] Zhong W, Zhao W (2021) Fetal ECG extraction using short time Fourier transform and generative adversarial networks. *Physiological Measurement*, 42(10):105011.

[290] Mohebbian MR, Vedaei SS, Wahid KA, Dinh A, Marateb HR, Tavakolian K (2022) Fetal ECG Extraction From Maternal ECG Using Attention-Based CycleGAN. *IEEE Journal of Biomedical and Health Informatics*, 26(2):515–526. https://doi.org/10.1109/JBHI.2021.3111873

[291] Luo Y, Chen Z, Yoshioka T (2020) Dual-Path RNN: Efficient Long Sequence Modeling for Time-Domain Single-Channel Speech Separation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2020-May:46–50. https://doi.org/10.1109/ICASSP40776.2020.9054266

[292] Andreotti F, Behar J, Clifford GD (2016) Fetal ECG Synthetic Database. https://doi.org/10.13026/C21P4T

[293] Matonia A, Jezewski J, Kupka T, Jezewski M, Horoba K, Wrobel J, Czabanski R, Kahankowa R (2020) Fetal electrocardiograms, direct and abdominal with reference heartbeat annotations. *Scientific data*, 7(1):200.

[294] Jezewski J, Matonia A, Kupka T, Roj D, Czabanski R (2012) Determination of fetal heart rate from abdominal signals: evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram. *Biomedizinische Technik/Biomedical Engineering*, 57(5):383–394.

[295] Andreotti F, Behar J, Clifford GD (2012) Abdominal and Direct Fetal ECG Database. https://doi.org/10.13026/C2RP4B

[296] Silva I, Behar J, Sameni R, Zhu T, Oster J, Clifford GD, Moody GB (2013) Noninvasive fetal ECG: the PhysioNet/computing in cardiology challenge 2013. *Computing in cardiology 2013*, :149–152.

[297] Behar J, Oster J, Clifford GD (2014) Combining and benchmarking methods offoetal ECG extraction without maternal or scalp electrode data. *Physiological Measurement*, 35(8):1569–1589. https://doi.org/10.1088/0967-3334/35/8/1569

[298] Behar J, Oster J, Clifford GD (2013) Non-invasive FECG extraction from a set of abdominal sensors.

[299] Shokouhmand A, Tavassolian N (2023) Fetal Electrocardiogram Extraction Using Dual-Path Source Separation of Single-Channel Non-Invasive Abdominal Recordings. *IEEE Transactions on Biomedical Engineering*, 70(1):283–295. https://doi.org/10.1109/TBME.2022.3189617

[300] Andreotti F, Gräßer F, Malberg H, Zaunseder S (2017) Non-invasive Fetal ECG Signal

Quality Assessment for Multichannel Heart Rate Estimation. *IEEE Transactions on Biomedical Engineering*, 64(12):2793–2802. https://doi.org/10.1109/TBME.2017.2675543

[301] Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: Lstm cells and network architectures. *Neural Computation*, 31(7):1235–1270. https://doi.org/10.1162/neco_a_01199

[302] Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116.

[303] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. *Advances in Neural Information Processing Systems*, 3(January):2672–2680. https://doi.org/10.3156/jsoft.29.5_177_2

[304] Singh P, Pradhan G (2021) A New ECG Denoising Framework Using Generative Adversarial Network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(2):759–764. https://doi.org/10.1109/TCBB.2020.2976981

[305] Kong J, Kim J, Bae J (2020) HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis. *Advances in Neural Information Processing Systems*, 33:17022–17033. https://proceedings.neurips.cc/paper/2020/file/c5d736809766d46260d816d8dbc9eb44-Paper.pdf

[306] Kolbæk M, Yu D, Tan Z-H, Jensen J (2017) Multitalker speech separation with utterance-level permutation invariant training of deep recurrent neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(10):1901–1913.

[307] Dietterich T (1995) Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, 27(3):326–327.

[308] Ying X (2019) An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168(2):22022.

[309] Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48.

[310] Iwana BK, Uchida S (2021) An empirical survey of data augmentation for time series classification with neural networks. *PLoS ONE*, 16(7 July):e0254841.

https://doi.org/10.1371/journal.pone.0254841

[311] Takahashi R, Matsubara T, Uehara K (2019) Data augmentation using random image cropping and patching for deep CNNs. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9):2917–2931.

[312] Mahdiani S, Jeyhani V, Peltokangas M, Vehkaoja A (2015) Is 50 Hz high enough ECG sampling frequency for accurate HRV analysis? *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015-Novem:5948–5951. https://doi.org/10.1109/EMBC.2015.7319746

[313] Ashrafimoghari V (2023) Detecting Cross-Lingual Information Gaps in Wikipedia. *Companion Proceedings of the ACM Web Conference 2023*, :581–585.

[314] Maaten L Van Der, Hinton G (2008) Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11):2579–2625.

ProQuest Number: 30568233

INFORMATION TO ALL USERS
The quality and completeness of this reproduction is dependent on the quality
and completeness of the copy made available to ProQuest.

Distributed by ProQuest LLC (2023).
Copyright of the Dissertation is held by the Author unless otherwise noted.

This work may be used in accordance with the terms of the Creative Commons license
or other rights statement, as indicated in the copyright statement or in the metadata
associated with this work. Unless otherwise specified in the copyright statement
or the metadata, all rights are reserved by the copyright holder.

This work is protected against unauthorized copying under Title 17,
United States Code and other applicable copyright laws.

Microform Edition where available © ProQuest LLC. No reproduction or digitization
of the Microform Edition is authorized without permission of ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346 USA